

Bayesian Inference in Nonparametric Dynamic State-Space Models

Anurag Ghosh, Soumalya Mukhopadhyay, Sandipan Roy and Sourabh Bhattacharya*

Abstract

We introduce state-space models where the functionals of the observational and the evolutionary equations are unknown, and treated as random functions evolving with time. Thus, our model is nonparametric and generalizes the traditional parametric state-space models. This random function approach also frees us from the restrictive assumption that the functional forms, although time-dependent, are of fixed forms. The traditional approach of assuming known, parametric functional forms is questionable, particularly in state-space models, since the validation of the assumptions require data on both the observed time series and the latent states; however, data on the latter are not available in state-space models.

We specify Gaussian processes as priors of the random functions and exploit the “look-up table approach” of Bhattacharya (2007) to efficiently handle the dynamic structure of the model. We consider both univariate and multivariate situations, using the Markov chain Monte Carlo (MCMC) approach for studying the posterior distributions of interest. In the case of challenging multivariate situations we demonstrate that the newly developed Transformation-based MCMC (TMCMC) of Dutta & Bhattacharya (2011) provides interesting and efficient alternatives to the usual proposal distributions. We illustrate our methods with a challenging multivariate simulated data set, where the true observational and the evolutionary equations are highly non-linear, and treated as unknown. The results we obtain are quite encouraging. Moreover, using our Gaussian process approach we analysed a real data set, which has also been

*Anurag Ghosh is a PhD student in Department of Statistical Science, Duke University, Soumalya Mukhopadhyay is a PhD student in Agricultural and Ecological Research Unit, Indian Statistical Institute, Sandipan Roy is a PhD student in Department of Statistics, University of Michigan, Ann Arbor, and Sourabh Bhattacharya is an Assistant Professor in Bayesian and Interdisciplinary Research Unit, Indian Statistical Institute, 203, B. T. Road, Kolkata 700108. Corresponding e-mail: sourabh@isical.ac.in.

analysed by Shumway & Stoffer (1982) and Carlin, Polson & Stoffer (1992) using the linearity assumption. Interestingly, the results of analyses of this data using our methods, although agree with those obtained by the previous authors for a major part of the time series, disagree towards the end of the time series. Since our methods very flexibly encapsulate both linear and nonlinear structures unlike the linear approaches of the previous authors, the disagreement between the results show that towards the end of the time series, the linearity assumption of the previous authors breaks down.

Keywords: *Evolutionary equation; Gaussian process; Look-up table; Observational equation; State-space model; Transformation-based Markov chain Monte Carlo.*

1. INTRODUCTION

The state-space models play important role in dealing with dynamic systems that arise in various disciplines such as finance, engineering, ecology, medicine, and statistics. The time-varying regression structure and the flexibility inherent in the sequential nature of state-space models make them very suitable for analysis and prediction of dynamic data. Indeed, as is well-known, most time series models of interest are expressible as state-space models; see Durbin & Koopman (2001) and Shumway & Stoffer (2011) for details. However, till date, the state-space models have considered only known forms of the equations, typically linear. But testing the parametric assumptions require data on both the observed time series and the unobserved states; unfortunately, data on the latter are not available in state-space models. Moreover, the regression structures of the state-space models may evolve with time, changing from linear to non-linear, and even the non-linear structure may also evolve with time, yielding further different non-linear structures. To our knowledge, there does not exist any approach in the literature that can handle such evolving, yet unknown, functional forms. These arguments point towards the need for developing general, nonparametric, approaches to state-space models, and this indeed, is our aim in this article. We adopt the Bayesian paradigm for its inherent flexibility.

In a nutshell, in this work, adopting a nonparametric Bayesian framework, we treat the regression structures as unknown and model these as Gaussian processes, and develop the consequent

theory in the Bayesian framework, considering both univariate and multivariate situations. The Gaussian process-based approach we develop here allows very flexible modeling of the unknown structures dynamically evolving with time, and gives particular importance to nonparametric dynamic evolutions of the state-space models, providing a sound and very general theory for the latter. We also develop efficient MCMC-based methods for simulating from the resulting posterior distributions.

Application of our methods to a challenging simulated multivariate data set with highly nonlinear true observational and evolutionary equations (treated as unknown) yielded encouraging results. Application of our ideas to a real data set which has been analysed by Shumway & Stoffer (1982) and Carlin et al. (1992) assuming linearity, provided an interesting insight that, although the linearity assumption may not be unreasonable for most part of the time series, linearity does not hold towards the end of the time series, which the previous authors failed to take account of. Indeed, the results we obtained here using our nonparametric dynamic state-space model are in agreement with the results obtained by the previous authors for a major part of the time series, but towards the end of the time series, our results disagree with those obtained using the linear assumptions. Since our nonparametric model includes both linear and nonlinear structures, the discrepancy between the results suggests that the linearity assumptions are no longer valid here after a certain point of time. These studies also vindicate that our approach is indeed capable of modeling unknown functions even if the forms are changing with time, without requiring any change point analysis and specification of functional forms before and after change points.

Before introducing our approach, we provide a brief overview of state-space models.

2. OVERVIEW OF STATE-SPACE MODELS

Generally, state-space models are of the following form: for $t = 1, 2, \dots$,

$$y_t = f_t(x_t) + \epsilon_t \tag{1}$$

$$x_t = g_t(x_{t-1}) + \eta_t \tag{2}$$

In the above, f_t and g_t are assumed to be functions of known forms which may or may not explicitly depend upon t ; η_t, ϵ_t are usually assumed to be zero mean *iid* normal variates. The choice $f_t(x_t) = F_t x_t$ and $g_t(x_{t-1}) = G_t x_{t-1}$, assuming known F_t, G_t , have found very wide use in the literature. Obviously, x_t, y_t may be univariate or multivariate. Matrix-variate dynamic linear models have been considered by Quintana & West (1987) and West & Harrison (1997) (see also Carvalho & West (2007)). Equation (1) is called the observational equation, while (2) is known as the evolutionary equation. Letting $\mathbf{D}_T = (y_1, \dots, y_T)'$ denote the available data, the goal is to obtain inferences about y_{T+1} (single-step forecast), y_{T+k} (k -step forecast), x_{T+1} conditional on y_{T+1} (filtering), x_{T-k} (retrospection). In the Bayesian paradigm, the interests center upon analyzing the corresponding posteriors $[y_{T+1} \mid \mathbf{D}_T]$, $[y_{T+k} \mid \mathbf{D}_T]$, $[x_{T+1} \mid \mathbf{D}_T, y_{T+1}]$ (also, $[x_{T+1} \mid \mathbf{D}_T]$) and $[x_{T-k} \mid \mathbf{D}_T]$.

In the non-Bayesian framework, solutions to dynamic systems are quite generally available via the well-known Kalman filter. However, the performance of Kalman filter is heavily dependent on the assumption of Gaussian errors and linearity of the functions in the observation and the evolution equations. In the case of non-linear dynamic models, various linearization techniques are used to obtain approximate solutions. For details on these issues, see Brockwell & Davis (1987), Meinhold & Singpurwala (1989), West & Harrison (1997) and the references therein. The Bayesian paradigm frees the investigator from restrictions of linear functions or Gaussian errors, and allows for very general dynamic model building through coherent combination of prior and the available time series data, and using Markov chain Monte Carlo (MCMC) for inference. Bayesian non-linear dynamic models with non-Gaussian errors, in conjunction with the Gibbs sampling approach for inference, have been considered in Carlin et al. (1992); see Durbin & Koopman (2001) and Shumway & Stoffer (2011) for comprehensive treatments of this area.

We are not aware of any nonparametric approach to modeling the functions f_t and g_t in state-space models. One can anticipate spline/wavelet-based approaches but such approaches are not capable of modeling the unknown functions when the functional forms change with time.

In the next section we introduce our novel nonparametric dynamic model where we use Gaussian processes to model the unknown functions f_t and g_t . Assuming the functions to be random

allows us to accommodate even those functions the forms of which are changing with time. We begin with one-dimensional x_t and y_t , gradually generalizing to the multivariate setup in Section 5. In Section 6 we illustrate our models and methodologies using a multivariate simulated data, where the true multivariate observational and the evolutionary equations are highly nonlinear, apart from being multidimensional. In Section 7 we consider application to a real, univariate data set. We present a summary of the current work, along with discussion of further work in Section 8. Additional derivations and further details are provided in the supplement Ghosh, Mukhopadhyay, Roy & Bhattacharya (2011b), whose sections have the prefix “S-” when referred to in this paper.

3. NONPARAMETRIC DYNAMIC MODEL: UNIVARIATE CASE

For any input x we represent $f_t(x)$ and $g_t(x)$ as $f(t, x)$ and $g(t, x)$, respectively. In other words, we consider time t as input as well. We denote $(t, x)'$ by x^* . Crucially, we allow $f(\cdot, \cdot)$ and $g(\cdot, \cdot)$ to be of unknown functional forms, which we model as Gaussian processes. In the univariate situation (the multivariate case will be considered subsequently) our complete dynamic model has the following form

$$y_t = f(x_{t,t}^*) + \epsilon_t, \quad \epsilon_t \sim N(0, \sigma_\epsilon^2), \quad (3)$$

$$x_t = g(x_{t,t-1}^*) + \eta_t, \quad \eta_t \sim N(0, \sigma_\eta^2), \quad (4)$$

where $x_{t,t}^* = (t, x_t)'$, and $x_{t,t-1}^* = (t, x_{t-1})'$. We assume that $x_0 \sim N(\mu_{x_0}, \sigma_{x_0}^2)$; $\mu_{x_0}, \sigma_{x_0}^2$ being known. The functions f and g are modeled as independent Gaussian processes with mean functions $\mu_f(\cdot) = \mathbf{h}(\cdot)' \boldsymbol{\beta}_f$ and $\mu_g(\cdot) = \mathbf{h}(\cdot)' \boldsymbol{\beta}_g$ with $\mathbf{h}(x^*) = (1, x^*)'$ for any x^* , and covariance functions of the form $\sigma_f^2 c_f(\cdot, \cdot)$ and $\sigma_g^2 c_g(\cdot, \cdot)$, respectively. The process variances are σ_f^2 and σ_g^2 and c_f, c_g are the correlation functions. Typically, for any $\mathbf{z}_1, \mathbf{z}_2$, $c_f(\mathbf{z}_1, \mathbf{z}_2) = \exp\{-(\mathbf{z}_1 - \mathbf{z}_2)' \mathbf{R}_f (\mathbf{z}_1 - \mathbf{z}_2)\}$ and $c_g(\mathbf{z}_1, \mathbf{z}_2) = \exp\{-(\mathbf{z}_1 - \mathbf{z}_2)' \mathbf{R}_g (\mathbf{z}_1 - \mathbf{z}_2)\}$, where \mathbf{R}_f and \mathbf{R}_g are 2×2 -dimensional diagonal matrices consisting of respective smoothness (or, roughness) parameters $\{r_{1,f}, r_{2,f}\}$ and $\{r_{1,g}, r_{2,g}\}$, which are responsible for the smoothness of the process realizations. These choices of the correlation functions imply that the functions, modeled by the process realizations, are infinitely smooth. The sets of parameters $\boldsymbol{\theta}_f = (\boldsymbol{\beta}_f, \sigma_f^2, \mathbf{R}_f)$ and $\boldsymbol{\theta}_g = (\boldsymbol{\beta}_g, \sigma_g^2, \mathbf{R}_g)$ are assumed

to be independent *a priori*. We consider the following form of prior distribution of the parameters:

$$[\boldsymbol{\beta}_f, \sigma_f^2, \mathbf{R}_f, \boldsymbol{\beta}_g, \sigma_g^2, r_g, \sigma_\epsilon^2, \sigma_\eta^2] = [\boldsymbol{\beta}_f, \sigma_f^2, \mathbf{R}_f][\boldsymbol{\beta}_g, \sigma_g^2, \mathbf{R}_g][\sigma_\epsilon^2, \sigma_\eta^2].$$

Note that, the distribution of \mathbf{D}_T , conditional on $x_{1,1}^*, \dots, x_{T,T}^*$ (equivalently, conditional on x_1, \dots, x_T), and the other parameters is multivariate normal:

$$\mathbf{D}_T \sim N_T(\mathbf{H}_{D_T} \boldsymbol{\beta}_f, \sigma_f^2 \mathbf{A}_{f,D_T} + \sigma_\epsilon^2 \mathbf{I}_T), \quad (5)$$

where $\mathbf{H}_{D_T}' = [\mathbf{h}(x_{1,1}^*), \dots, \mathbf{h}(x_{T,T}^*)]$, \mathbf{A}_{f,D_T} is a $T \times T$ matrix with (i, j) -th element $c_f(x_{i,i}^*, x_{j,j}^*)$; $(i, j) = 1, \dots, T$, and \mathbf{I}_T is the T -th order identity matrix.

The joint distribution of (x_1, \dots, x_T) , however, is much less straightforward. Observe that, although we have $[x_0] \sim N(\mu_{x_0}, \sigma_{x_0}^2)$, $[x_1 | x_0] \sim N(\mathbf{h}(x_0)' \boldsymbol{\beta}_g, \sigma_g^2 + \sigma_\eta^2)$, but $[x_2 | x_1, x_0] = [g(2, x_1) + \eta_2 | x_1, x_0] = [g(2, g(1, x_0) + \eta_1) + \eta_2 | g(1, x_0) + \eta_1, x_0]$, the rightmost expression suggesting that special techniques may be necessary to get hold of the conditional distribution. We adopt the procedure introduced by Bhattacharya (2007) to deal with this problem. The idea is to conceptually simulate the entire function g modeled by the Gaussian process, and use the simulated process as a look-up table to obtain the conditional distributions of $\{x_i; i \geq 2\}$.

3.1 Look-up table idea for facilitating derivation of conditionals and ensuring computational efficiency

For the purpose of illustration only let us assume that $\epsilon_t = 0$ for all t , yielding the model $x_t = g(x_{t,t-1}^*)$. The concept of look-up table in this problem can be briefly explained as follows. Given that the entire process $g(\cdot)$ is available (this means that for every input z , the corresponding $g(z)$ is available), conditional on $x_{t,t-1}^*$, $x_t = g(x_{t,t-1}^*)$ can be obtained by simply picking input $x_{t,t-1}^*$ and reporting the corresponding output value $g(x_{t,t-1}^*)$. This hypothetical simulation procedure suggests that conditional on the simulated process, it can be safely assumed that x_t depends only upon $x_{t,t-1}^*$.

Note that given x_0 we can simulate $x_1 = g(x_{1,0}^*) \sim N(\mathbf{h}(x_{1,0}^*)' \boldsymbol{\beta}_g, \sigma_g^2)$, which is the marginal distribution of the Gaussian process prior. Thus, x_1 is simulated without resorting to any approximation. It then remains to simulate the rest of the dynamic sequence, for which we need to simulate

the rest of the process $\{g(x^*); x^* \neq x_{1,0}^*\}$. In practice, it is not possible to have a simulation of this entire set. We only have available a set of grid points $\mathbf{G}_z = \{z_1, \dots, z_n\}$ obtained, perhaps, by Latin hypercube sampling (see, for example, Santner, Williams & Notz (2003)) and a corresponding simulation of g , given by $\mathbf{D}_z = \{g(z_1), \dots, g(z_n)\}$, the latter having a joint multivariate normal distribution with mean

$$E[\mathbf{D}_z \mid \boldsymbol{\theta}_g] = \mathbf{H}_{D_z} \boldsymbol{\beta}_g \quad (6)$$

and covariance matrix

$$V[\mathbf{D}_z \mid \boldsymbol{\theta}_g] = \sigma_g^2 \mathbf{A}_{g,D_z}, \quad (7)$$

where $\mathbf{H}'_{D_z} = [\mathbf{h}(z_1), \dots, \mathbf{h}(z_n)]$ and \mathbf{A}_{g,D_z} is a correlation matrix with the (i, j) -th element $c_g(z_i, z_j)$. and $\mathbf{s}_{g,D_z}(\cdot) = (c_g(\cdot, z_1), \dots, c_g(\cdot, z_n))'$.

Given $(x_0, g(x_{1,0}^*))$, we simulate \mathbf{D}_z from $[\mathbf{D}_z \mid \boldsymbol{\theta}_g, g(x_{1,0}^*), x_0]$. Since the joint distribution of $[\mathbf{D}_z, g(x_{1,0}^*) \mid x_0]$ is multivariate normal with mean vector $\begin{pmatrix} \mathbf{H}_{D_z} \boldsymbol{\beta}_g \\ \mathbf{h}(x_0)' \boldsymbol{\beta}_g \end{pmatrix}$ and covariance matrix $\sigma_g^2 \mathbf{A}_{D_z, g(x_{1,0}^*)}$ where

$$\mathbf{A}_{D_z, g(x_{1,0}^*)} = \begin{pmatrix} \mathbf{A}_{g,D_z} & \mathbf{s}_{g,D_z}(x_{1,0}^*) \\ \mathbf{s}_{g,D_z}(x_{1,0}^*)' & 1 \end{pmatrix}, \quad (8)$$

it follows that the conditional $[\mathbf{D}_z \mid g(x_{1,0}^*), x_{1,0}^*]$ has an n -variate normal distribution with mean vector

$$E[\mathbf{D}_z \mid g(x_{1,0}^*), x_0, \boldsymbol{\theta}_g] = \boldsymbol{\mu}_{g,D_z} = \mathbf{H}_{D_z} \boldsymbol{\beta}_g + \mathbf{s}_{g,D_z}(x_{1,0}^*)(g(x_{1,0}^*) - \mathbf{h}(x_{1,0}^*)' \boldsymbol{\beta}_g) \quad (9)$$

and covariance matrix

$$V[\mathbf{D}_z \mid g(x_{1,0}^*), x_0, \boldsymbol{\theta}_g] = \sigma_g^2 \boldsymbol{\Sigma}_{g,D_z} \quad (10)$$

where

$$\boldsymbol{\Sigma}_{g,D_z} = \mathbf{A}_{g,D_z} - \mathbf{s}_{g,D_z}(x_{1,0}^*) \mathbf{s}_{g,D_z}(x_{1,0}^*)' \quad (11)$$

The probability that $x_{t,t-1}^*$ will coincide with some grid point in \mathbf{G}_z , is zero. But the Gaussian process assumption provides a nice way of dealing with the problem. Assuming conditional independence as discussed above, the conditional distribution $[g(x_{t,t-1}^*) \mid \mathbf{D}_z, x_{t-1}]$ is normal with

mean

$$\mu_t = \mathbf{h}(x_{t,t-1}^*)' \boldsymbol{\beta}_g + \mathbf{s}_{g,D_z}(x_{t,t-1}^*)' \mathbf{A}_{g,D_z}^{-1} (\mathbf{D}_z - \mathbf{H}_{D_z} \boldsymbol{\beta}_g) \quad (12)$$

and variance

$$\sigma_t^2 = \sigma_g^2 \{1 - \mathbf{s}_{g,D_z}(x_{t,t-1}^*)' \mathbf{A}_{g,D_z}^{-1} \mathbf{s}_{g,D_z}(x_{t,t-1}^*)\} \quad (13)$$

Note that, thanks to the Gaussian process assumption, conditioning on \mathbf{D}_z forces the random function $g(\cdot)$ to pass through the points in $(\mathbf{G}_z, \mathbf{D}_z)$ since the conditional $[g(x) \mid \boldsymbol{\theta}_g, \mathbf{D}_z, x]$ has zero variance if $x \in \mathbf{G}_z$ (see, for example, Bhattacharya (2007) and the references therein). In other words, if $x_{t,t-1}^* \in \mathbf{G}_z$, then $\sigma_t^2 = 0$ so that

$$[g(x_{t,t-1}^*) \mid \boldsymbol{\theta}_g, \mathbf{D}_z, x_{t-1}] = \delta_{g(x_{t,t-1}^*)}, \quad (14)$$

where δ_z denotes point mass at z . This property of the conditional associated with Gaussian process is in keeping with the insight gained from the discussion related to look-up table associated with prediction of the outputs of deterministic function having dynamic behaviour. However, $x_{t,t-1}^* \notin \mathbf{G}_z$ with probability 1 and the conditional $[g(x_{t,t-1}^*) \mid \boldsymbol{\theta}_g, \mathbf{D}_z, x_{t-1}]$ provides spatial interpolation within $(\mathbf{G}_z, \mathbf{D}_z)$ (see, for example, Cressie (1993), Stein (1999)). Finer the set \mathbf{G}_z , closer is $[g(x_{t,t-1}^*) \mid \boldsymbol{\theta}_g, \mathbf{D}_z, x_{t-1}]$ to $\delta_{g(x_{t,t-1}^*)}$. The conditional independence assumption of $g(x_{t,t-1}^*)$ of all $\{x_{t,i}^*; i < (t-1)\}$ given $(\mathbf{G}_z, \mathbf{D}_z)$ is in accordance with the motivation provided by the deterministic sequence and here \mathbf{D}_z acts as a set of auxiliary variables, greatly simplifying computation, while not compromising on accuracy. One subtlety involved in the assumption of conditional independence is that, conditional on $x_{1,0}^*$, \mathbf{G}_z must not contain $x_{1,0}^*$; otherwise \mathbf{D}_z would contain $g(x_{1,0}^*)$ implying that $[g(x_{t,t}^*) \mid \boldsymbol{\theta}_g, \mathbf{D}_z, x_t]$ is dependent on $x_1 = g(x_{1,0}^*)$, violating the conditional independence assumption.

To summarize the ideas, let $x_0 \sim \pi$, where π is some appropriate prior distribution of x_0 . The entire dynamic sequence $\{x_0, x_1 = g(x_{1,0}^*), x_2 = g(x_{2,1}^*), \dots\}$ can then be simulated using the following steps sequentially:

- (1) Draw $x_0 \sim \pi$.
- (2) Given x_0 , draw $x_1 = g(x_{1,0}^*) \sim N(\mathbf{h}(x_{1,0}^*)' \boldsymbol{\beta}_g, \sigma_g^2)$.

(3) Given x_0 , and $x_1 = g(x_{1,0}^*)$, draw $\mathbf{D}_z \sim [\mathbf{D}_z \mid \boldsymbol{\theta}_g, g(x_{1,0}^*), x_0]$.

(4) For $t = 2, 3, \dots$, draw $x_t \sim [x_t = g(x_{t,t-1}^*) \mid \boldsymbol{\theta}_g, \mathbf{D}_z, x_{t-1}]$.

Step (1) is a simulation of x_0 from its prior, step (2) is simply drawn from the known marginal distribution of $g(x_{1,0}^*)$ given x_0 . In step (3) \mathbf{D}_z is drawn conditional on $g(x_{1,0}^*)$ (and $x_{1,0}^*$), conceptually implying that the rest of the process $\{g(x^*); x \neq x_{1,0}^*\}$ is drawn once $g(x_{1,0}^*)$ is known. Step (4) then uses this simulated \mathbf{D}_z to obtain the rest of the dynamic sequence, using the assumed conditional independence structure.

It is shown in Bhattacharya (2007), marginalization over \mathbf{D}_z is possible but the resulting computations become severely unstable numerically. In our state space modeling scenario, although we use the concept of look-up table illustrated above, the posterior distributions are much more complex and analytic marginalization over \mathbf{D}_z is not possible, since the posterior of \mathbf{D}_z is intractable. In any case, numerical instability is avoided in our situation.

3.2 Look-up table-induced joint distribution of $\{x_0, x_1, \dots, x_{T+1}, \mathbf{D}_z\}$

Once \mathbf{G}_z and \mathbf{D}_z are available, we write down the joint distribution of $(\mathbf{D}_z, x_0, x_1, \dots, x_T)$ conditional on the other parameters as

$$\begin{aligned} & [x_0, x_1, \dots, x_T, \mathbf{D}_z \mid \boldsymbol{\theta}_f, \boldsymbol{\theta}_g, \sigma_\epsilon^2, \sigma_\eta^2] \\ &= [x_0][x_1 = g(x_{1,0}^*) + \eta_1 \mid x_0, \sigma_\eta^2][\mathbf{D}_z \mid \boldsymbol{\theta}_g] \\ & \quad \prod_{t=1}^T [x_{t+1} = g(x_{t+1,t}^*) + \eta_{t+1} \mid \mathbf{D}_z, x_t, \boldsymbol{\theta}_g, \sigma_\eta^2] \end{aligned} \quad (15)$$

Recall that $[x_0] \sim N(\mu_{x_0}, \sigma_{x_0}^2)$, $[x_1 = g(x_{1,0}^*) + \eta_1 \mid x_{1,0}^*, \sigma_\eta^2] \sim N(\mathbf{h}(x_{1,0}^*)'\boldsymbol{\beta}_g, \sigma_g^2 + \sigma_\eta^2)$ and the distribution of \mathbf{D}_z is multivariate normal with mean and variance given by (6) and (7). The conditional distribution $[x_{t+1} = g(x_{t+1,t}^*) + \eta_{t+1} \mid \mathbf{D}_z, x_t, \boldsymbol{\theta}_g, \sigma_\eta^2]$ is normal with mean

$$\mu_{x_t} = \mathbf{h}(x_{t+1,t}^*)'\boldsymbol{\beta}_g + \mathbf{s}_{g,\mathbf{D}_z}(x_{t+1,t}^*)'\mathbf{A}_{g,\mathbf{D}_z}^{-1}(\mathbf{D}_z - \mathbf{H}_{\mathbf{D}_z}\boldsymbol{\beta}_g) \quad (16)$$

and variance

$$\sigma_{x_t}^2 = \sigma_\eta^2 + \sigma_g^2 \{1 - \mathbf{s}_{g,\mathbf{D}_z}(x_{t+1,t}^*)'\mathbf{A}_{g,\mathbf{D}_z}^{-1}\mathbf{s}_{g,\mathbf{D}_z}(x_{t+1,t}^*)\} \quad (17)$$

Observe that in this case even if $x_{t+1,t}^* \in \mathbf{G}_z$, due to the presence of the additive error term η_{t+1} , the conditional variance of x_{t+1} is non-zero, equalling $\sigma_{x_t}^2 = \sigma_\eta^2$, the error variance. Conditional on all the unknowns, the distribution of \mathbf{D}_T is given by (5). In the next section we complete specification of our fully Bayesian model by choosing appropriate prior distributions of the parameters.

3.3 Prior specifications

We assume the following prior distributions:

$$[\log(r_{i,f})] \sim N\left(\mu_{r_{i,f}}, \sigma_{r_{i,f}}^2\right); \text{ for } i = 1, 2. \quad (18)$$

$$[\log(r_{i,g})] \sim N\left(\mu_{r_{i,g}}, \sigma_{r_{i,g}}^2\right); \text{ for } i = 1, 2. \quad (19)$$

$$[\sigma_\epsilon^2] \propto (\sigma_\epsilon^2)^{-\left(\frac{\alpha_\epsilon+2}{2}\right)} \exp\left\{-\frac{\gamma_\epsilon}{2\sigma_\epsilon^2}\right\}; \alpha_\epsilon, \gamma_\epsilon > 0 \quad (20)$$

$$[\sigma_\eta^2] \propto (\sigma_\eta^2)^{-\left(\frac{\alpha_\eta+2}{2}\right)} \exp\left\{-\frac{\gamma_\eta}{2\sigma_\eta^2}\right\}; \alpha_\eta, \gamma_\eta > 0 \quad (21)$$

$$[\sigma_f^2] \propto (\sigma_f^2)^{-\left(\frac{\alpha_f+2}{2}\right)} \exp\left\{-\frac{\gamma_f}{2\sigma_f^2}\right\}; \alpha_f, \gamma_f > 0 \quad (22)$$

$$[\sigma_g^2] \propto (\sigma_g^2)^{-\left(\frac{\alpha_g+2}{2}\right)} \exp\left\{-\frac{\gamma_g}{2\sigma_g^2}\right\}; \alpha_g, \gamma_g > 0 \quad (23)$$

$$[\boldsymbol{\beta}_f] \sim N_m\left(\boldsymbol{\beta}_{f,0}, \boldsymbol{\Sigma}_{\beta_{f,0}}\right) \quad (24)$$

$$[\boldsymbol{\beta}_g] \sim N_m\left(\boldsymbol{\beta}_{g,0}, \boldsymbol{\Sigma}_{\beta_{g,0}}\right) \quad (25)$$

All the prior parameters are assumed to be known. We discuss the choices of the prior parameters in the contexts of the specific applications.

3.4 MCMC-based Bayesian inference

In this section we begin with the problem of forecasting y_{T+1} , given the data set \mathbf{D}_T . Interestingly, our approach to this problem provides an MCMC methodology which generates inference about all the posterior distributions required, either as by-products or by simple generalization of this MCMC approach using an augmentation scheme. Details follow.

The posterior predictive distribution of y_{T+1} given \mathbf{D}_T is

$$\begin{aligned} [y_{T+1} \mid \mathbf{D}_T] &= \int [y_{T+1} \mid \mathbf{D}_T, x_0, x_1, \dots, x_{T+1}, \boldsymbol{\beta}_f, \mathbf{R}_f, \sigma_\epsilon^2] \\ &\quad \times [x_0, x_1, \dots, x_{T+1}, \boldsymbol{\beta}_f, \boldsymbol{\beta}_g, \mathbf{R}_f, \mathbf{R}_g, \sigma_\epsilon^2 \sigma_\eta^2 \mid \mathbf{D}_T] d\boldsymbol{\theta}_f d\boldsymbol{\theta}_g d\sigma_\epsilon^2 d\sigma_\eta^2 dx_0 dx_1, \dots, dx_{t+1}. \end{aligned} \quad (26)$$

This posterior is not available analytically, and so simulation methods are necessary to make inferences. In particular, once a sample is available from the posterior $[x_0, x_1, \dots, x_{t+1}, \boldsymbol{\beta}_f, \boldsymbol{\beta}_g, \mathbf{R}_f, \mathbf{R}_g, \sigma_\epsilon^2 \sigma_\eta^2 \mid \mathbf{D}_T]$, the corresponding samples drawn from $[y_{T+1} \mid \mathbf{D}_T, x_0, x_1, \dots, x_{T+1}, \boldsymbol{\beta}_f, \mathbf{R}_f, \sigma_\epsilon^2]$ are from the posterior predictive (26), using which required posterior summaries can be obtained. Note that the conditional distribution $[y_{T+1} = f(x_{T+1,T+1}^*) + \epsilon_{T+1} \mid \mathbf{D}_T, x_0, x_1, \dots, x_{T+1}, \boldsymbol{\beta}_f, \sigma_f^2, \mathbf{R}_f, \sigma_\epsilon^2]$ is normal with mean

$$\mu_{y_{T+1}} = \mathbf{h}(x_{T+1,T+1}^*)' \boldsymbol{\beta}_f + \mathbf{s}_{f,\mathbf{D}_T}(x_{T+1,T+1}^*)' \mathbf{A}_{f,\mathbf{D}_T}^{-1} (\mathbf{D}_T - \mathbf{H}_{\mathbf{D}_T} \boldsymbol{\beta}_f) \quad (27)$$

and variance

$$\sigma_{y_{T+1}}^2 = \sigma_\epsilon^2 + \sigma_f^2 \{1 - \mathbf{s}_{f,\mathbf{D}_T}(x_{T+1,T+1}^*)' \mathbf{A}_{f,\mathbf{D}_T}^{-1} \mathbf{s}_{f,\mathbf{D}_T}(x_{T+1,T+1}^*)\} \quad (28)$$

Using \mathbf{D}_z , the conditional posterior $[x_0, x_1, \dots, x_{T+1}, \boldsymbol{\beta}_f, \boldsymbol{\beta}_g, \mathbf{R}_f, \mathbf{R}_g, \sigma_\epsilon^2, \sigma_\eta^2 \mid \mathbf{D}_T]$ can be written as

$$[x_0, x_1, \dots, x_{T+1}, \boldsymbol{\theta}_f, \boldsymbol{\theta}_g, \sigma_\epsilon^2, \sigma_\eta^2 \mid \mathbf{D}_T] = \int [x_0, x_1, \dots, x_{T+1}, \boldsymbol{\theta}_f, \boldsymbol{\theta}_g, \sigma_\epsilon^2, \sigma_\eta^2, g(x_{1,0}^*), \mathbf{D}_z \mid \mathbf{D}_T] dg(x_{1,0}^*) d\mathbf{D}_z \quad (29)$$

$$\propto \int [x_0, x_1, \dots, x_{T+1}, \boldsymbol{\theta}_f, \boldsymbol{\theta}_g, \sigma_\epsilon^2, \sigma_\eta^2, g(x_{1,0}^*), \mathbf{D}_z, \mathbf{D}_T] dg(x_0) d\mathbf{D}_z \quad (30)$$

$$\begin{aligned} &= \int [\boldsymbol{\theta}_f, \boldsymbol{\theta}_g, \sigma_\epsilon^2, \sigma_\eta^2][x_0][g(x_{1,0}^*) \mid x_{1,0}^*][\mathbf{D}_z \mid g(x_{1,0}^*), x_0, \boldsymbol{\theta}_g] \\ &\quad \times [x_1 = g(x_{1,0}^*) + \eta_1 \mid g(x_{1,0}^*), x_0, \boldsymbol{\beta}_g, \sigma_g^2, \sigma_\eta^2] \\ &\quad \times [x_{T+1} = g(x_{T+1,T}^*) + \eta_T \mid \boldsymbol{\theta}_g, \sigma_\eta^2, \mathbf{D}_z, x_T] \\ &\quad \times \prod_{t=1}^{T-1} [x_{t+1} \mid \boldsymbol{\theta}_g, \sigma_\eta^2, \mathbf{D}_z, x_t][\mathbf{D}_T \mid x_1, \dots, x_T, \boldsymbol{\theta}_f, \sigma_\epsilon^2] dg(x_{1,0}^*) d\mathbf{D}_z \end{aligned} \quad (31)$$

In the above, $[x_1 = g(x_{1,0}^*) + \eta_1 \mid g(x_{1,0}^*), \beta_g, \sigma_g^2, \sigma_\eta^2] \sim N(g(x_{1,0}^*), \sigma_\eta^2)$. Although the analytic form of (31) is not available, MCMC simulation from $[x_0, x_1, \dots, x_{T+1}, \theta_f, \theta_g, \sigma_\epsilon^2, \sigma_\eta^2, g(x_{1,0}^*), D_z \mid D_T]$, which is proportional to the integrand in (31), is possible. Ignoring $g(x_{1,0}^*)$ and D_z in these MCMC simulations yields the desired samples from $[x_0, x_1, \dots, x_{T+1}, \theta_f, \theta_g, \sigma_\epsilon^2, \sigma_\eta^2 \mid D_T]$. In the next section we provide the forms of the full conditionals necessary for MCMC simulations. The complete details are provided in Section S-9.

4. FORMS OF THE FULL CONDITIONAL DISTRIBUTIONS FOR MCMC SAMPLING IN THE UNIVARIATE SITUATION

Let $G_{z0} = G_z \cup \{x_{1,0}^*\}$, $D_{z0} = (D'_z, g(x_{1,0}^*))'$, $A_{g,z0} = \begin{pmatrix} A_{g,D_z} & s_{g,D_z}(x_{1,0}^*) \\ s_{g,D_z}(x_{1,0}^*)' & 1 \end{pmatrix}$ and $H'_{g,D_{z0}} = [H'_{g,D_z}, h(x_{1,0}^*)]$. Clearly, $[D_{z0} \mid x_{1,0}^*, \theta_g] = [g(x_{1,0}^*) \mid x_0, \theta_g][D_z \mid g(x_{1,0}^*), x_0, \theta_g]$.

With these definitions, the forms of the full conditional distributions of the unknowns are provided

below.

$$[\boldsymbol{\beta}_f \mid \cdots] \propto [\boldsymbol{\beta}_f][\boldsymbol{D}_T \mid x_1, \dots, x_T, \boldsymbol{\theta}_f, \sigma_\epsilon^2] \quad (32)$$

$$[\boldsymbol{\beta}_g \mid \cdots] \propto [\boldsymbol{\beta}_g][\boldsymbol{D}_{z0} \mid x_{1,0}^*, \boldsymbol{\theta}_g] \prod_{t=1}^T [x_{t+1} \mid \boldsymbol{\theta}_g, \sigma_\eta^2, \boldsymbol{D}_z, x_t] \quad (33)$$

$$[\sigma_f^2 \mid \cdots] \propto [\sigma_f^2][\boldsymbol{D}_T \mid x_1, \dots, x_T, \boldsymbol{\theta}_f, \sigma_\epsilon^2] \quad (34)$$

$$[\sigma_g^2 \mid \cdots] \propto [\sigma_g^2][\boldsymbol{D}_{z0} \mid x_{1,0}^*, \boldsymbol{\theta}_g] \prod_{t=1}^T [x_{t+1} \mid \boldsymbol{\theta}_g, \sigma_\eta^2, \boldsymbol{D}_z, x_t] \quad (35)$$

$$[\sigma_\epsilon^2 \mid \cdots] \propto [\sigma_\epsilon^2][\boldsymbol{D}_T \mid x_1, \dots, x_T, \boldsymbol{\theta}_f, \sigma_\epsilon^2] \quad (36)$$

$$[\sigma_\eta^2 \mid \cdots] \propto [\sigma_\eta^2][x_1 \mid g(x_{1,0}^*), x_0, \boldsymbol{\theta}_g, \sigma_\eta^2] \prod_{t=1}^T [x_{t+1} \mid \boldsymbol{\theta}_g, \sigma_\eta^2, \boldsymbol{D}_z, x_t] \quad (37)$$

$$[r_{i,f} \mid \cdots] \propto [r_{i,f}][\boldsymbol{D}_T \mid x_1, \dots, x_T, \boldsymbol{\theta}_f, \sigma_\epsilon^2]; \quad i = 1, 2 \quad (38)$$

$$[r_{i,g} \mid \cdots] \propto [r_{i,g}][\boldsymbol{D}_z \mid g(x_{1,0}^*), x_0, \boldsymbol{\theta}_g] \prod_{t=1}^T [x_{t+1} \mid \boldsymbol{\theta}_g, \sigma_\eta^2, \boldsymbol{D}_z, x_t]; \quad i = 1, 2 \quad (39)$$

$$[g(x_{1,0}^*) \mid \cdots] \propto [g(x_{1,0}^*) \mid x_0, \boldsymbol{\beta}_g, \sigma_g^2][\boldsymbol{D}_z \mid g(x_{1,0}^*), x_0, \boldsymbol{\theta}_g][x_1 \mid g(x_{1,0}^*), x_0, \sigma_\eta^2] \quad (40)$$

$$[\boldsymbol{D}_z \mid \cdots] \propto \prod_{t=1}^T [x_{t+1} \mid \boldsymbol{\theta}_g, \sigma_\eta^2, \boldsymbol{D}_z, x_t][\boldsymbol{D}_z \mid g(x_{1,0}^*), x_0, \boldsymbol{\theta}_g] \quad (41)$$

$$[x_0 \mid \cdots] \propto [x_0][\boldsymbol{D}_{z0} \mid x_0, \boldsymbol{\theta}_g] \quad (42)$$

$$[x_1 \mid \cdots] \propto [x_1 \mid g(x_{1,0}^*), x_0, \boldsymbol{\beta}_g, \sigma_g^2, \sigma_\eta^2][x_2 \mid \boldsymbol{\theta}_g, \sigma_\eta^2, \boldsymbol{D}_z, x_1] \\ \times [\boldsymbol{D}_T \mid x_1, \dots, x_T, \boldsymbol{\theta}_f, \sigma_\epsilon^2] \quad (43)$$

$$[x_{t+1} \mid \cdots] \propto [x_{t+1} \mid \boldsymbol{\theta}_g, \sigma_\eta^2, \boldsymbol{D}_z, x_t][x_{t+2} \mid \boldsymbol{\theta}_g, \sigma_\eta^2, \boldsymbol{D}_z, x_{t+1}] \\ \times [\boldsymbol{D}_T \mid x_1, \dots, x_T, \boldsymbol{\theta}_f, \sigma_\epsilon^2]; \quad t = 1, \dots, T-1 \quad (44)$$

$$[x_{T+1} \mid \cdots] \propto [x_{T+1} \mid \boldsymbol{\theta}_g, \sigma_\eta^2, \boldsymbol{D}_z, x_T] \quad (45)$$

Although some of the full conditionals are of standard forms permitting Gibbs sampling steps, others are non-standard and sampling requires Metropolis-Hastings (MH) steps in those situations. In Section S-9, we describe the Gibbs steps and construct proposal distributions when MH steps are needed.

4.0.1. MCMC-based single and mutiple step forecasts, filtering and retrospection Observe that one can readily study the posteriors $[y_{T+1} \mid \mathbf{D}_T]$, $[x_{T+1} \mid \mathbf{D}_T]$ and $[x_{T-k} \mid \mathbf{D}_T]$ using the readily available MCMC samples of $\{y_{T+1}, x_{T+1}, x_{T-k}\}$ after ignoring the samples corresponding to the rest of the unknowns. To study the posterior $[x_{T+1} \mid \mathbf{D}_T, y_{T+1}]$, we only need to augment y_{T+1} to \mathbf{D}_T to create $\mathbf{D}_{T+1} = (\mathbf{D}'_T, y_{T+1})'$. Then our methodology can be followed exactly to generate samples from $[x_{T+1} \mid \mathbf{D}_{T+1}]$. Sample generation from $[y_{T+k} \mid \mathbf{D}_T]$ requires a slight generalization of this augmentation strategy. Here we use successive augmentation, adding each simulated y_{T+j} to the previous \mathbf{D}_{T+j-1} to create $\mathbf{D}_{T+j} = (\mathbf{D}'_{T+j-1}, y_{T+j})$; $j = 1, 2, \dots, k$. Then our MCMC methodology can be implemented successively to generate samples from y_{T+j+1} and all other variables. This implies that at each augmentation stage we need to draw a single MCMC sample from $[x_0, x_1, \dots, x_{T+j+1}, \boldsymbol{\beta}_f, \boldsymbol{\beta}_g, \mathbf{R}_f, \mathbf{R}_g, \sigma_f^2, \sigma_g^2, \sigma_\epsilon^2, \sigma_\eta^2 \mid \mathbf{D}_{T+j}]$. Once this sample is generated, we can draw a single realization from y_{T+j+1} by drawing from $y_{T+j+1} \sim N\left(\mu_{y_{T+j+1}}, \sigma_{y_{T+j+1}}^2\right)$, where, analogous to (27) and (28),

$$\mu_{y_{T+j+1}} = \mathbf{h}(x_{T+j+1, T+j+1}^*)' \boldsymbol{\beta}_f + \mathbf{s}_{f, \mathbf{D}_{T+j}}(x_{T+j+1, T+j+1}^*)' \mathbf{A}_{f, \mathbf{D}_{T+j}}^{-1} (\mathbf{D}_{T+j} - \mathbf{H}_{\mathbf{D}_{T+j}} \boldsymbol{\beta}_f) \quad (46)$$

and variance

$$\sigma_{y_{T+j+1}}^2 = \sigma_\epsilon^2 + \sigma_f^2 \left\{ 1 - \mathbf{s}_{f, \mathbf{D}_{T+j}}(x_{T+j+1, T+j+1}^*)' \mathbf{A}_{f, \mathbf{D}_{T+j}}^{-1} \mathbf{s}_{f, \mathbf{D}_{T+j}}(x_{T+j+1, T+j+1}^*) \right\}. \quad (47)$$

5. EXTENSION TO MULTIVARIATE SITUATION

Now we extend our nonparametric dynamic model to the case where both \mathbf{y}_t and \mathbf{x}_t are multivariate. In particular, we assume that they are p -component and q -component vectors, respectively. Then our multivariate model is of the form

$$\mathbf{y}_t = \mathbf{f}(\mathbf{x}_{t,t}^*) + \boldsymbol{\epsilon}_t, \quad \boldsymbol{\epsilon}_t \sim N_p(\mathbf{0}, \boldsymbol{\Sigma}_\epsilon), \quad (48)$$

$$\mathbf{x}_t = \mathbf{g}(\mathbf{x}_{t,t-1}^*) + \boldsymbol{\eta}_t, \quad \boldsymbol{\eta}_t \sim N_q(\mathbf{x}_0, \boldsymbol{\Sigma}_\eta), \quad (49)$$

where $\mathbf{x}_0 \sim N_q(\boldsymbol{\mu}_{x_0}, \boldsymbol{\Sigma}_{x_0})$; $\boldsymbol{\mu}_{x_0}, \boldsymbol{\Sigma}_{x_0}$ assumed known. In the above, $\mathbf{f}(\cdot) = (f_1(\cdot), \dots, f_p(\cdot))'$ is a function with p components and $\mathbf{g}(\cdot) = (g_1(\cdot), \dots, g_q(\cdot))'$ is a function consisting of q components. We assume that $\mathbf{f}(\cdot)$ is a p -variate Gaussian process with mean $E[\mathbf{f}(\cdot)] = \mathbf{B}'_f \mathbf{h}(\cdot)$ and

with covariance function $cov(\mathbf{f}(\mathbf{z}_1), \mathbf{f}(\mathbf{z}_2)) = c_f(\mathbf{z}_1, \mathbf{z}_2)\mathbf{\Sigma}_f$, for any q -dimensional inputs $\mathbf{z}_1, \mathbf{z}_2$. Here $\mathbf{h}(\cdot) = (h_1(\cdot), \dots, h_m(\cdot))'$ and $\mathbf{B}_f = (\boldsymbol{\beta}_{1,f}, \dots, \boldsymbol{\beta}_{p,f})$, where, for $j = 1, \dots, p$, $\boldsymbol{\beta}_{j,f}$ are m -dimensional column vectors; clearly, $m = q+2$. Also, $c_f(\mathbf{z}_1, \mathbf{z}_2) = \exp\{-(\mathbf{z}_1 - \mathbf{z}_2)' \mathbf{R}_f (\mathbf{z}_1 - \mathbf{z}_2)\}$, where \mathbf{R}_f is a diagonal matrix consisting of $(q+1)$ smoothness parameters, denoted by $\{r_{1,f}, \dots, r_{(q+1),f}\}$. Similarly, we assume that $\mathbf{g}(\cdot)$ is a Gaussian process with mean $E[\mathbf{g}(\cdot)] = \mathbf{B}'_g \mathbf{h}(\cdot)$ and covariance function $c_g(\mathbf{z}_1, \mathbf{z}_2)\mathbf{\Sigma}_g = \exp\{-(\mathbf{z}_1 - \mathbf{z}_2)' \mathbf{R}_g (\mathbf{z}_1 - \mathbf{z}_2)\}$, the notation used being analogous to those used for description of the Gaussian process $\mathbf{f}(\cdot)$.

5.1 Multivariate data and its distribution

The multivariate data is given by the following $T \times p$ matrix: $\mathbf{D}_T = (\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_T)'$. Writing \mathbf{D}_T vectorically as a Tp -vector for convenience, as $\mathbf{D}_{Tp} = (\mathbf{y}'_1, \mathbf{y}'_2, \dots, \mathbf{y}'_T)'$, it follows that $[\mathbf{D}_{Tp} \mid \mathbf{B}_f, \mathbf{\Sigma}_f, \mathbf{R}_f, \mathbf{\Sigma}_\epsilon]$ is a Tp -variate normal with mean vector

$$E[\mathbf{D}_{Tp} \mid \mathbf{B}_f, \mathbf{\Sigma}_f, \mathbf{R}_f, \mathbf{\Sigma}_\epsilon] = \begin{pmatrix} \mathbf{B}'_f \mathbf{h}(\mathbf{x}_{1,1}^*) \\ \mathbf{B}'_f \mathbf{h}(\mathbf{x}_{2,2}^*) \\ \vdots \\ \mathbf{B}'_f \mathbf{h}(\mathbf{x}_{T,T}^*) \end{pmatrix} = \boldsymbol{\mu}_{D_{Tp}} \text{ (say)} \quad (50)$$

and covariance matrix

$$V[\mathbf{D}_{Tp} \mid \mathbf{B}_f, \mathbf{\Sigma}_f, \mathbf{R}_f, \mathbf{\Sigma}_\epsilon] = \mathbf{A}_{f,D_T} \otimes \mathbf{\Sigma}_f + \mathbf{I}_T \otimes \mathbf{\Sigma}_\epsilon = \mathbf{\Sigma}_{D_{Tp}} \text{ (say)}. \quad (51)$$

Since (51) does not admit a Kronecker product form, the distribution of \mathbf{D}_{Tp} can not be written as matrix normal, which requires right and left covariance matrices forming Kronecker product in the corresponding multivariate distribution of the vector. It follows that $[\mathbf{y}_{T+1} = \mathbf{f}(\mathbf{x}_{T+1,T+1}^*) + \boldsymbol{\epsilon}_{T+1} \mid \mathbf{D}_T, \mathbf{x}_0, \mathbf{x}_1, \dots, \mathbf{x}_{T+1} \mid \mathbf{B}_f, \mathbf{\Sigma}_f, \mathbf{R}_f, \mathbf{\Sigma}_\epsilon]$ is p -variate normal with mean

$$\boldsymbol{\mu}_{y_{T+1}} = \mathbf{B}'_f \mathbf{h}(\mathbf{x}_{T+1,T+1}^*) + (\mathbf{D}_T - \mathbf{H}_{D_T} \mathbf{B}_f)' \mathbf{A}_{f,D_T}^{-1} \mathbf{s}_{f,D_T}(\mathbf{x}_{T+1,T+1}^*) \quad (52)$$

and variance

$$\mathbf{\Sigma}_{y_{T+1}} = \{1 - \mathbf{s}_{f,D_T}(\mathbf{x}_{T+1,T+1}^*)' \mathbf{A}_{f,D_T}^{-1} \mathbf{s}_{f,D_T}(\mathbf{x}_{T+1,T+1}^*)\} \mathbf{\Sigma}_f + \mathbf{\Sigma}_\epsilon \quad (53)$$

5.2 Distributions of $\mathbf{g}(\mathbf{x}_{1,0}^*)$ and \mathbf{D}_z

Conditional on \mathbf{x}_0 , $\mathbf{g}(\mathbf{x}_{1,0}^*)$ is q -variate normal with mean $\mathbf{B}_g' \mathbf{h}(\mathbf{x}_{1,0}^*)$ and covariance matrix Σ_g .

In contrast with the distribution of \mathbf{D}_T , $\mathbf{D}_{z,nq} = (\mathbf{g}'(\mathbf{z}_1), \mathbf{g}'(\mathbf{z}_2), \dots, \mathbf{g}'(\mathbf{z}_n))'$ has an nq -variate normal distribution with mean

$$E[\mathbf{D}_{z,nq} \mid \mathbf{B}_g, \Sigma_g, \mathbf{R}_g] = \begin{pmatrix} \mathbf{B}_g' \mathbf{h}(\mathbf{z}_1) \\ \mathbf{B}_g' \mathbf{h}(\mathbf{z}_2) \\ \vdots \\ \mathbf{B}_g' \mathbf{h}(\mathbf{z}_n) \end{pmatrix} = \boldsymbol{\mu}_{D_{z,nq}} \quad (\text{say}) \quad (54)$$

and covariance matrix

$$V[\mathbf{D}_{z,nq} \mid \mathbf{B}_g, \Sigma_g, \mathbf{R}_g] = \mathbf{A}_{g,D_z} \otimes \Sigma_g = \Sigma_{D_{z,nq}} \quad (\text{say}). \quad (55)$$

Hence, the distribution of the $n \times q$ -dimensional matrix $\mathbf{D}_z = (\mathbf{g}(\mathbf{z}_1), \mathbf{g}(\mathbf{z}_2), \dots, \mathbf{g}(\mathbf{z}_n))'$ is matrix normal:

$$[\mathbf{D}_z \mid \mathbf{B}_g, \Sigma_g, \mathbf{R}_g] \sim \mathcal{N}_{n,q}(\mathbf{H}_z \mathbf{B}_g, \mathbf{A}_{g,D_z}, \Sigma_g) \quad (56)$$

Conditionally on $(\mathbf{x}_0, \mathbf{g}(\mathbf{x}_{1,0}^*))$, it follows that \mathbf{D}_z is $n \times q$ -dimensional matrix-normal:

$$[\mathbf{D}_z \mid \mathbf{g}(\mathbf{x}_{1,0}^*), \mathbf{x}_0, \mathbf{B}_g, \Sigma_g, \mathbf{R}_g, \Sigma_\eta] \sim \mathcal{N}_{n,q}(\boldsymbol{\mu}_{g,D_z}, \Sigma_{g,D_z}, \Sigma_g) \quad (57)$$

In (57) $\boldsymbol{\mu}_{g,D_z}$ is the mean matrix, given by

$$\boldsymbol{\mu}_{g,D_z} = \mathbf{H}_{D_z} \mathbf{B}_g + \mathbf{s}_{g,D_z}(\mathbf{x}_{1,0}^*)(\mathbf{g}(\mathbf{x}_{1,0}^*)' - \mathbf{h}(\mathbf{x}_{1,0}^*)' \mathbf{B}_g), \quad (58)$$

and

$$\Sigma_{g,D_z} = \mathbf{A}_{g,D_z} - \mathbf{s}_{g,D_z}(\mathbf{x}_{1,0}^*) \mathbf{s}_{g,D_z}(\mathbf{x}_{1,0}^*)'. \quad (59)$$

Here we slightly abuse notation to denote both univariate and multivariate versions of the mean matrix and the right covariance matrix by $\boldsymbol{\mu}_{g,D_z}$ and Σ_{g,D_z} , respectively (see (9) and (11)).

5.3 Joint distribution of $\{\mathbf{x}_0, \dots, \mathbf{x}_{T+1}, \mathbf{D}_z\}$

Note that

$$[\mathbf{x}_1 \mid \mathbf{g}(\mathbf{x}_0), \mathbf{x}_0, \mathbf{B}_g, \Sigma_g] \sim N_q(\mathbf{g}(\mathbf{x}_{1,0}^*), \Sigma_\eta), \quad (60)$$

and for $t = 1, \dots, T$, the conditional distribution $[\mathbf{x}_{t+1} = \mathbf{g}(\mathbf{x}_{t+1,t}^*) + \boldsymbol{\eta}_{t+1} \mid \mathbf{D}_z, \mathbf{x}_t, \mathbf{B}_g, \boldsymbol{\Sigma}_g, \mathbf{R}_g, \boldsymbol{\Sigma}_\eta]$ is q -variate normal with mean

$$\boldsymbol{\mu}_{x_t} = \mathbf{B}_g' \mathbf{h}(\mathbf{x}_{t+1,t}^*) + (\mathbf{D}_z - \mathbf{H}_{D_z} \mathbf{B}_g)' \mathbf{A}_{g,D_z}^{-1} \mathbf{s}_{g,D_z}(\mathbf{x}_{t+1,t}^*) \quad (61)$$

and variance

$$\boldsymbol{\Sigma}_{x_t} = \{1 - \mathbf{s}_{g,D_z}(\mathbf{x}_{t+1,t}^*)' \mathbf{A}_{g,D_z}^{-1} \mathbf{s}_{g,D_z}(\mathbf{x}_{t+1,t}^*)\} \boldsymbol{\Sigma}_g + \boldsymbol{\Sigma}_\eta. \quad (62)$$

Since $[\mathbf{x}_0] \sim N_p(\boldsymbol{\mu}_{x_0}, \boldsymbol{\Sigma}_{x_0})$ and the distribution of \mathbf{D}_z is given by (56) the joint distribution is obtained by taking products of the individual distributions.

5.4 Prior distributions

We assume the following prior distributions:

For $i = 1, \dots, (q+1)$,

$$[\log(r_{i,f})] \stackrel{iid}{\sim} N(\mu_{R_f}, \sigma_{R_f}^2) \quad (63)$$

For $i = 1, \dots, (p+1)$,

$$[\log(r_{i,g})] \stackrel{iid}{\sim} N(\mu_{R_g}, \sigma_{R_g}^2) \quad (64)$$

$$[\boldsymbol{\Sigma}_\epsilon] \propto |\boldsymbol{\Sigma}_\epsilon|^{-\frac{\nu_\epsilon + p + 1}{2}} \exp \left[-\frac{1}{2} \text{tr}(\boldsymbol{\Sigma}_\epsilon^{-1} \boldsymbol{\Sigma}_{\epsilon,0}) \right]; \quad \nu_\epsilon > p - 1 \quad (65)$$

$$[\boldsymbol{\Sigma}_\eta] \propto |\boldsymbol{\Sigma}_\eta|^{-\frac{\nu_\eta + q + 1}{2}} \exp \left[-\frac{1}{2} \text{tr}(\boldsymbol{\Sigma}_\eta^{-1} \boldsymbol{\Sigma}_{\eta,0}) \right]; \quad \nu_\eta > q - 1 \quad (66)$$

$$[\boldsymbol{\Sigma}_f] \propto |\boldsymbol{\Sigma}_f|^{-\frac{\nu_f + p + 1}{2}} \exp \left[-\frac{1}{2} \text{tr}(\boldsymbol{\Sigma}_f^{-1} \boldsymbol{\Sigma}_{f,0}) \right]; \quad \nu_f > p - 1 \quad (67)$$

$$[\boldsymbol{\Sigma}_g] \propto |\boldsymbol{\Sigma}_g|^{-\frac{\nu_g + q + 1}{2}} \exp \left[-\frac{1}{2} \text{tr}(\boldsymbol{\Sigma}_g^{-1} \boldsymbol{\Sigma}_{g,0}) \right]; \quad \nu_g > q - 1 \quad (68)$$

$$[\mathbf{B}_f \mid \boldsymbol{\Sigma}_f] \sim \mathcal{N}_{m,p}(\mathbf{B}_{f,0}, \boldsymbol{\Sigma}_{B_f,0}, \psi \boldsymbol{\Sigma}_f) \quad (69)$$

$$[\mathbf{B}_g \mid \boldsymbol{\Sigma}_g] \sim \mathcal{N}_{m,q}(\mathbf{B}_{g,0}, \boldsymbol{\Sigma}_{B_g,0}, \psi \boldsymbol{\Sigma}_g) \quad (70)$$

All the prior parameters are assumed to be known; and their choices will be discussed in the specific applications to be considered.

The full conditionals are of analogous forms as those in the univariate case, provided by equations from (32) to (45), but now the univariate distributions must be replaced with multivariate

distributions and multivariate distributions with matrix-variate distributions, and interestingly in some cases the full conditionals are not available in closed form although these were available in the one-dimensional version of the problem. For example, although the full conditional of β_f and β_g were available in the one-dimensional problem, the corresponding distributions of \mathbf{B}_f and \mathbf{B}_g are no longer available in closed forms. The problem of obtaining the closed form of the full conditional of \mathbf{B}_f can be attributed to the fact that (51) can not be represented as a single Kronecker product. The non-availability of the closed form in the case of \mathbf{B}_g is due to fact that the covariance matrices of \mathbf{x}_t are additive. In fact, it turns out that only the full conditional of \mathbf{x}_{T+1} is of standard form. Details of our MCMC methods for the multivariate case are provided in Section S-10.

6. SIMULATION STUDY

We consider a simulation study where we generate the data from a 4-variate model, constructed using the univariate nonstationary growth model of Carlin et al. (1992). To reduce computational burden in this multi-dimensional set up we generated 50 data points using the following model: for $t = 1, \dots, 50$,

$$x_{t,1} = \alpha x_{t-1,1} + \beta x_{t-1,1} / (1 + x_{t-1,1}^2) + \gamma \cos(1.2(t-1)) + \eta_{t,1} \quad (71)$$

$$x_{t,2} = \alpha x_{t-1,2} + \beta x_{t-1,2} / (1 + x_{t-1,2}^2) + \eta_{t,2} \quad (72)$$

$$x_{t,3} = \alpha + \beta x_{t-1,3} + \eta_{t,3} \quad (73)$$

$$x_{t,4} = \gamma \cos(1.2(t-1)) + \eta_{t,4} \quad (74)$$

$$y_{t,i} = x_{t,i}^2 / 20 + \epsilon_{t,i}; \quad i = 1, \dots, 4, \quad (75)$$

where $\epsilon_{t,i}$ and $\eta_{t,i}$ are *iid* zero-mean normal random variables with variances 0.1. We set $\alpha = 0.05$, $\beta = 0.1$ and $\gamma = 0.2$, and $\mathbf{x}_0 = \mathbf{0}$.

6.1 Choice of prior parameters

For the priors of \mathbf{B}_f and \mathbf{B}_g , we set $\psi = 1$, $\mathbf{B}_{f,0}$ and $\mathbf{B}_{g,0}$ to be null matrices, and $\Sigma_{B_f,0}$ and $\Sigma_{B_g,0}$ to be identity matrices. For the priors of Σ_ϵ , Σ_η , Σ_f , and Σ_g we set $\nu_\epsilon = p$, $\nu_\eta = q$, $\nu_f = p$, and $\nu_g = q$. We chose $\Sigma_\epsilon = \Sigma_\eta = 0.1\mathbf{I}$, and $\Sigma_f = \Sigma_g = 0.5\mathbf{I}$, where \mathbf{I} denotes the identity matrix.

For the log-normal priors of the smoothness parameters we set $\mu_{R_f} = \mu_{R_g} = -0.5$ and $\sigma_{R_f}^2 = \sigma_{R_g}^2 = 1$. The choices imply that the prior mean and the prior variance of each of the smoothness parameters are, respectively, 1 and 2 (approximately). In our simulation experiment these choices seemed adequate. In the case of real data analysis, in Section 7, however, we increased the prior variabilities, since in real data cases more uncertainty about smoothness is expected.

6.2 MCMC implementation

For updating the smoothness parameters we used normal random walk proposals with variances 0.005. To set up the grid G_z for the model-fitting purpose, we considered $[-1.5, 1.5]^4$ to be a grid space for the 4-dimensional variable \mathbf{x} . Indeed, this grid-space contains the entire true 4-dimensional time series. We divide $[-1.5, 1.5]$ into 100 equal sub-intervals and chose a point randomly from each of 100 sub-intervals, in each dimension, yielding $n = 100$ 4-dimensional points corresponding to \mathbf{x} . As before, we chose the first component of the grid (corresponding to the time component) by uniformly drawing from each subinterval $[i, i + 1]; i = 0, \dots, 99$.

The block updating proposal for updating B_f , described in Section S-10.1. worked quite well, but those for block updating B_g , $g(\mathbf{x}_{1,0}^*)$, D_z , and those for the covariance matrices, as described in Section S-10. yielded poor acceptance rates. In order to update the covariance matrices Σ_ϵ , Σ_η , Σ_g and Σ_f , we considered the following strategy: we re-wrote the matrices in the form CC' , where C is a lower triangular matrix, and used normal random walk with variance 0.005, to update the non-zero elements in a single block. This improved the acceptance rates. For B_g , $g(\mathbf{x}_{1,0}^*)$ and D_z , the strategy of block updating using random walk failed. As a remedy we considered the transformation-based MCMC (TMCMC), recently developed by Dutta & Bhattacharya (2011); in particular, we used the additive transformation, which requires much less number of move types and hence computationally less expensive compared to other, non-additive move types. Briefly, for each of the blocks, we generated $\xi \sim N(0, 0.05)I(\xi > 0)$. Then, for each parameter in the block, we either added or subtracted ξ with equal probability. In other words, we used the same ξ to update all the parameters in a block, unlike the block random walk proposal. This considerably improved the acceptance rates. For theoretical and implementation details, see Dutta & Bhattacharya (2011).

We discarded the first 10,000 iterations of the MCMC run as burn-in and stored the subsequent 50,000 iterations for inference. Convergence is assessed by informal diagnostics, as before. It took an ordinary laptop about 24 hours to implement this multivariate experiment.

6.3 Results of model-fitting

Figures 1 and 2 show that our Gaussian process-based nonparametric model performs well in spite of the multidimensional situation—the true values are well-captured by the posterior distributions, and the true forecast values are also well-supported by the corresponding forecast distributions of \mathbf{x}_{51} and \mathbf{y}_{51} . Moreover, each component of the true 4-variate time series fall entirely within the 95% highest posterior densities of the corresponding component of the 4-variate posterior time series.

We now apply our ideas based on Gaussian processes to a real data set.

7. APPLICATION TO A REAL DATA SET

Assuming a parametric, dynamic, linear model set up, Shumway & Stoffer (1982) and Carlin et al. (1992) analysed a data set consisting of estimated total physician expenditures by year ($y_t; t = 1, \dots, 25$) as measured by the Social Security Administration. The unobserved true annual physician expenditures are denoted by $x_t; t = 1, \dots, 25$. Shumway & Stoffer (1982) used a maximum likelihood approach based on the EM algorithm, while Carlin et al. (1992) considered a Gibbs sampling based Bayesian approach.

We apply our Bayesian nonparametric Gaussian process based model and methodology on the same data set and check if the results of the former authors who analysed this data based on the linearity assumption, agree with those obtained by our far more general analysis.

7.1 Choice of prior parameters and MCMC implementation

For the choice of the parameters of the priors of σ_f^2 , σ_g^2 , σ_ϵ^2 and σ_η^2 , we first note that the mean is of the form $\gamma/(\alpha - 2)$ and the variance is of the form $2\gamma^2/\{(\alpha - 2)^2(\alpha - 4)\}$. Thus, if we set $\gamma/(\alpha - 2) = a$, then the variance becomes $2a^2/(\alpha - 4)$. Here we set $a = 0.5, 0.5, 0.1, 0.1$, respectively, for σ_f^2 , σ_g^2 , σ_ϵ^2 , σ_η^2 . We set $\alpha = 4.01$ so that the prior variance is of the form $200a^2$; we set

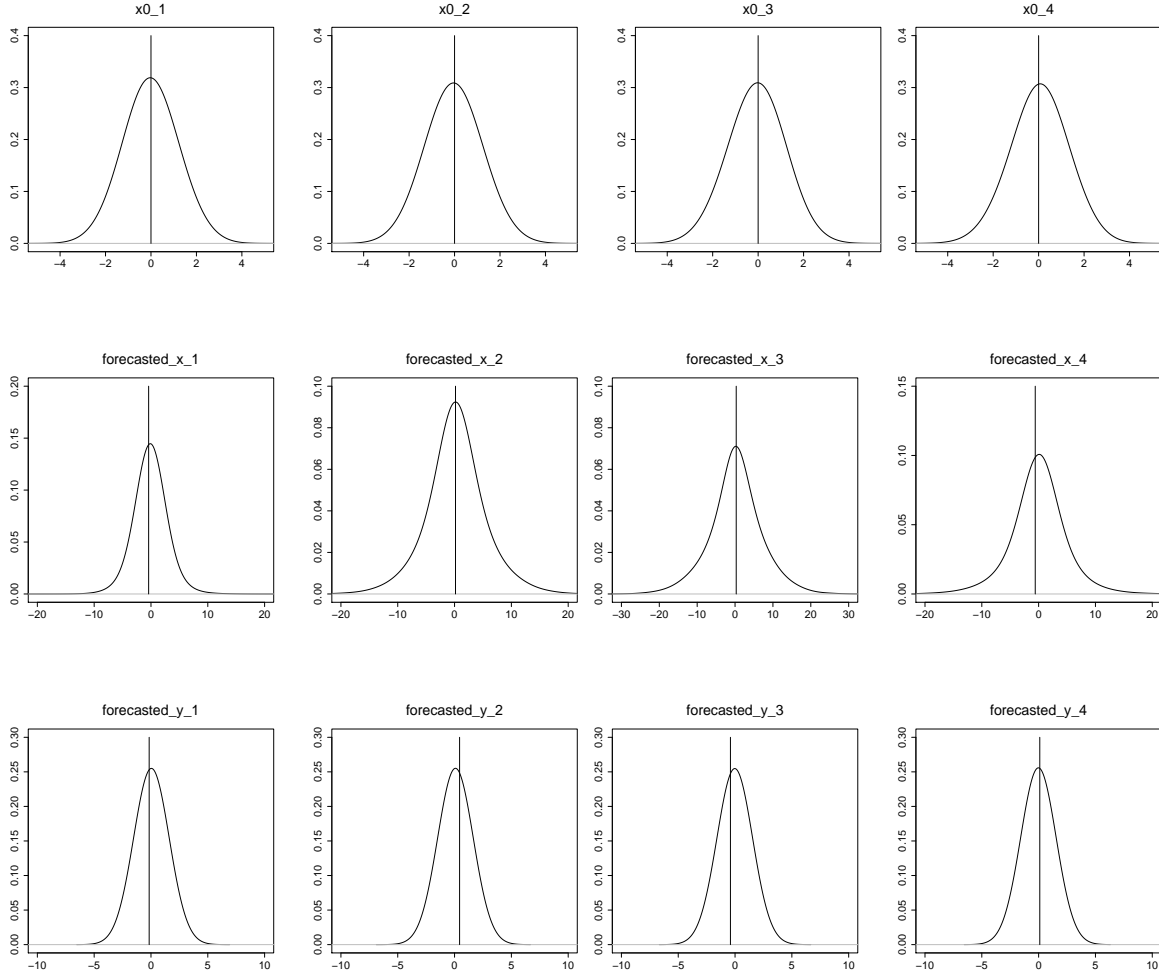


Figure 1: **Simulation study with data generated from the 4-variate variate modification of the model of CPS:** Posterior distributions of \mathbf{x}_0 , \mathbf{x}_{T+1} (one-step forecasted \mathbf{x}), and \mathbf{y}_{T+1} (one-step forecasted \mathbf{y}). The solid line stands for the true values of the respective parameters.

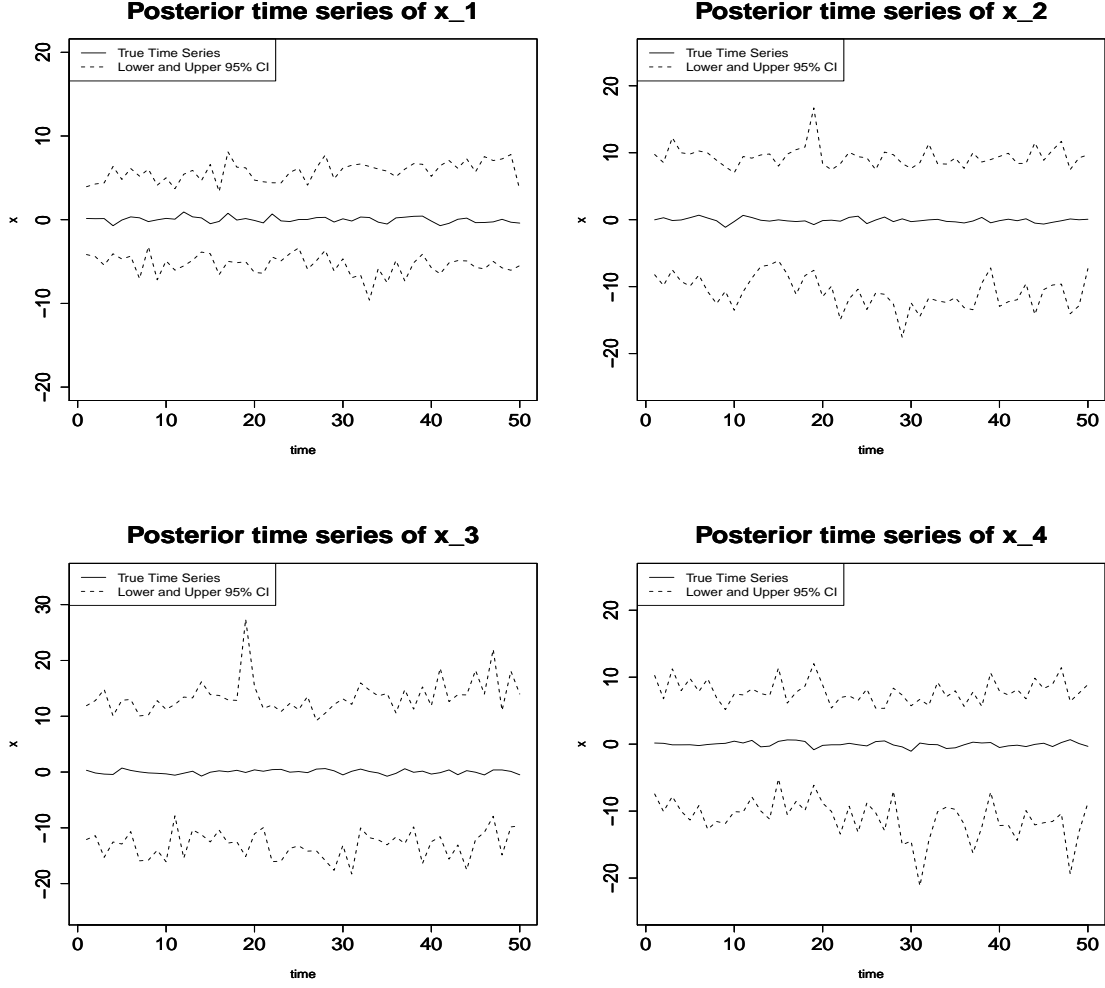


Figure 2: **Simulation study with data generated from the 4-variate modification of the model of CPS:** 95% highest posterior density credible intervals of the time series $x_{1,j}, \dots, x_{T,j}$; $j = 1, 2, 3, 4$. The solid line stands for the true time series.

$a = 0.5, 0.5, 100, 100$, respectively, for σ_f^2 , σ_g^2 , σ_ϵ^2 and σ_η^2 . These imply that the prior expectations of the variances are 0.5, 0.5, 100, and 100, respectively, along with the aforementioned variances. The choices of high values of the prior expectations of σ_ϵ^2 and σ_η^2 are motivated by Carlin et al. (1992), while the small variabilities of σ_f^2 and σ_g^2 reflects the belief that the true functional forms are perhaps not very different from linearity, the belief being motivated by the assumption of linearity by both Shumway & Stoffer (1982) and Carlin et al. (1992). Also motivated by the latter we chose $x_0 \sim N(2500, 100^2)$ to be the prior distribution of x_0 . We set the priors of β_f and β_g to be trivariate normal distributions with zero mean and the identity matrix as the covariance matrix. For the prior parameters of the smoothness parameters we set $\sigma^2 = 100$ and $\mu = -50$ in the log-normal prior distributions, so that the means are 1 and the variances are $\exp(100) - 1$. As mentioned in Section 6.1, here we set high variance for the smoothness parameters to account for much higher degree of uncertainty about smoothness in this real data situation compared to the simulation experiment.

To set up the grid G_z we noted that the MLEs of the time series obtained by Shumway & Stoffer (1982) using linear state space model are contained in $[2000, 30000]$. We divide this interval into 200 sub-intervals of equal length, and select a value randomly from each such sub-interval, obtaining values of the second component of the two-dimensional grid G_z . For the first component, we generate a number uniformly from each of the 200 sub-intervals $[i, i + 1]; i = 0, \dots, 199$.

We discard the first 10,000 MCMC iterations as burn-in and store the next 50,000 iterations for inference. We used the normal random walk proposal with variance 0.05 for updating σ_f , σ_g , σ_ϵ and σ_η and the normal random walk proposal with variance 10 for updating $\{x_0, x_1, \dots, x_T\}$. These choices of the variances are based on pilot runs of our MCMC algorithm. As before, informal diagnostics indicated good convergence properties of our MCMC algorithm. It took around 17 hours to implement this application in an ordinary laptop machine.

7.2 Results of model-fitting

Figures 3, 4 and 5 show the posterior distributions of the unknowns. Our posterior time series of x_t has relatively narrow 95% highest posterior density credible intervals, vindicating that the linearity

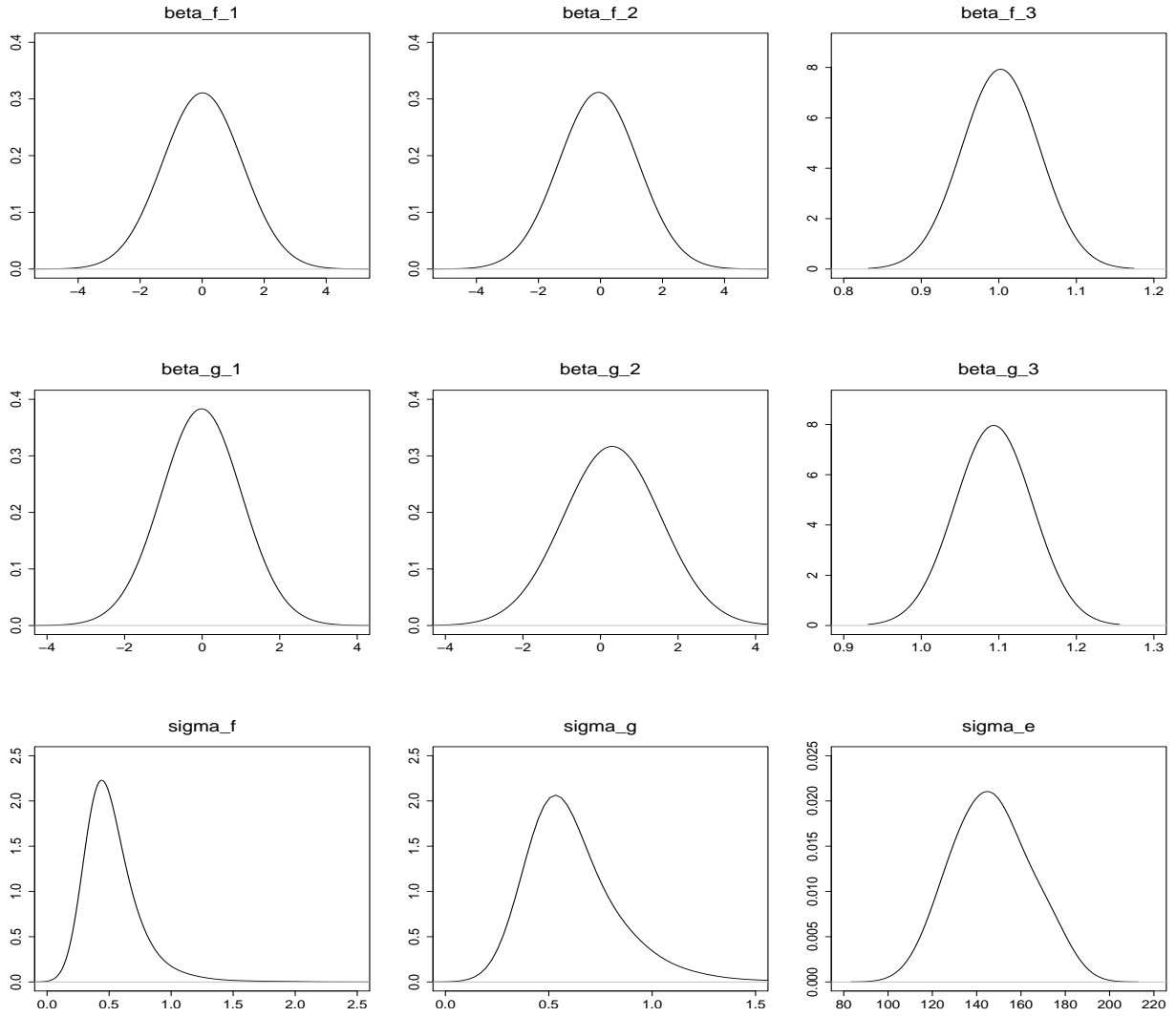


Figure 3: **Real data analysis:** Posterior densities of $\beta_{0,f}$, $\beta_{1,f}$, $\beta_{2,f}$, $\beta_{0,g}$, $\beta_{1,g}$, $\beta_{2,g}$, σ_f , σ_g , and σ_e .

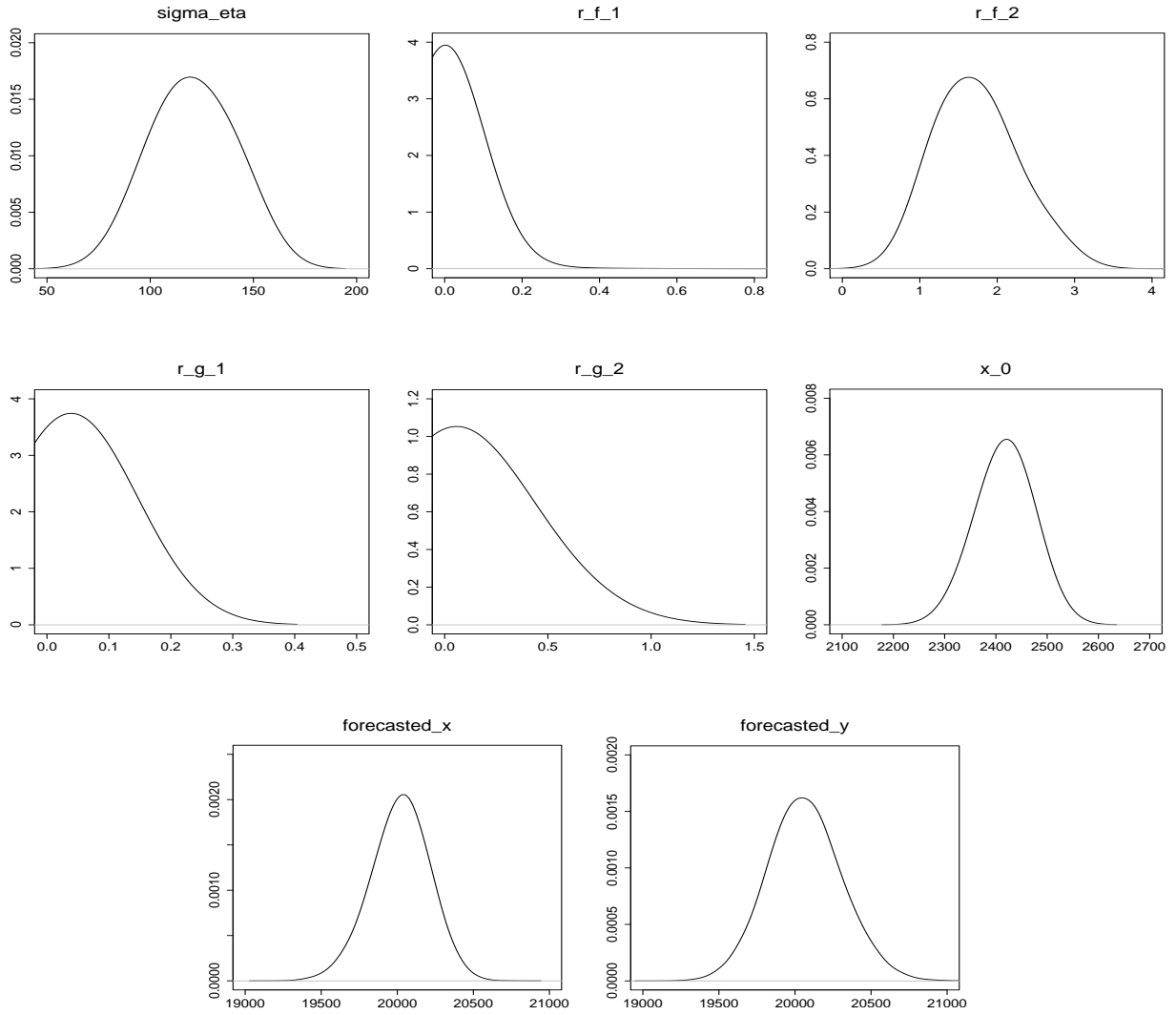


Figure 4: **Real data analysis:** Posterior densities of σ_η , $r_{1,f}$, $r_{2,f}$, $r_{1,g}$, $r_{2,g}$, x_0 , x_{T+1} (one-step forecasted x), and y_{T+1} (one-step forecasted y).

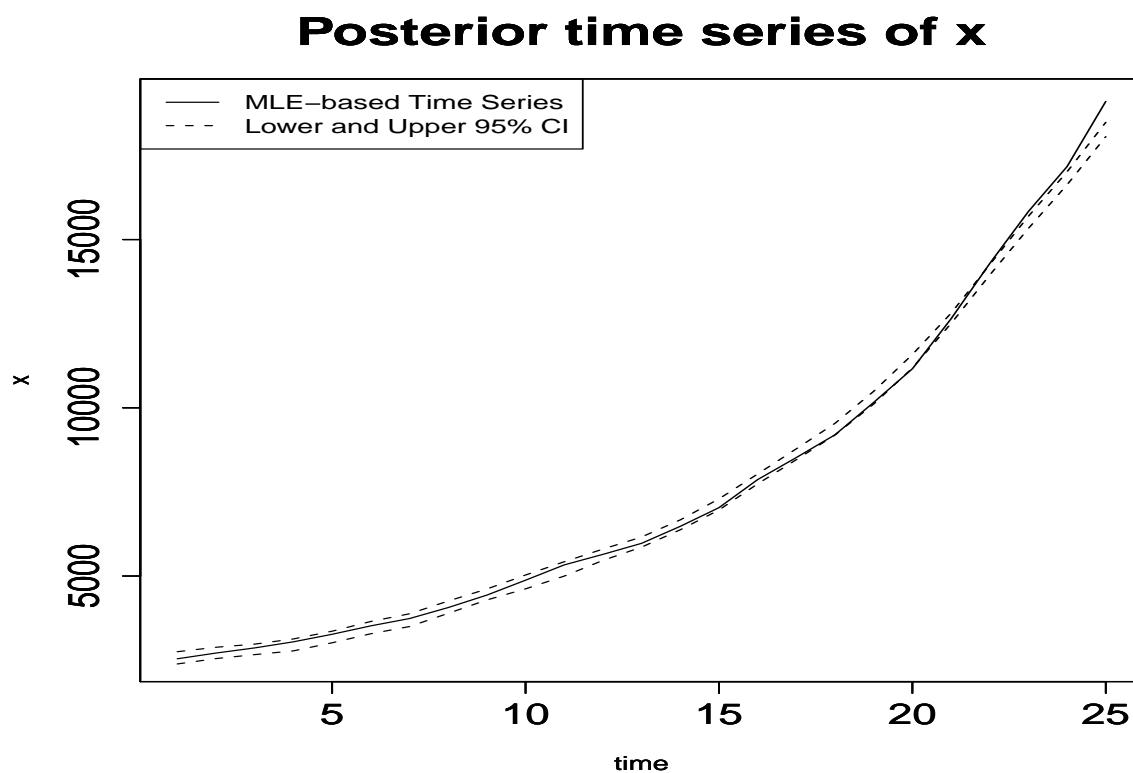


Figure 5: **Real data analysis:** 95% highest posterior density credible intervals of the time series x_1, \dots, x_T . The solid line stands for the MLE time series obtained by Shumway and Stoffer (1982).

assumption is not unreasonable as claimed by Shumway & Stoffer (1982) and Carlin et al. (1992). Indeed, for such linear relationships, it is well-known that the Gaussian process priors that we adopted are expected to perform very well.

However, perhaps not all is well with the aforementioned linearity assumption of Shumway & Stoffer (1982) and Carlin et al. (1992). Note that although the MLE time series obtained by Shumway & Stoffer (1982) fall mostly within our Bayesian 95% highest posterior density credible intervals of x_t , five observations towards the end of the time series fall outside. These observations correspond to the years 1968, 1970, 1971, 1972, and 1973. The values of y_t in these years are 11,099, 14,306, 15,835, 16,916, and 18,200. We suspect that linearity breaks down towards the end of the time series, which the models of Shumway & Stoffer (1982) and Carlin et al. (1992) fail to take account of. On the other hand, without any change point analysis, our flexible non-parametric model, based on Gaussian processes, is able to accommodate changes in the regression structures.

8. CONCLUSIONS AND FUTURE WORK

In this article, using Gaussian process priors and the “look-up table” idea of Bhattacharya (2007) we proposed a novel methodology for Bayesian inference in nonparametric state space models, in both univariate and multivariate cases. The Gaussian process priors on the unknown functional forms of the observational and the evolutionary equations allow for very flexible modeling of time-varying random functions, where even the functional forms may change over time, without requiring any change point analysis. We have vindicated the effectiveness of our model and methodology with a challenging multidimensional simulation experiment and a real data analysis which provided interesting insight into nonlinearity of the underlying time series towards its end.

For our current purpose, we have assumed *iid* Gaussian noises ϵ_t and η_t ; however, it is straightforward to generalize these to other parametric distributions (thick-tailed or otherwise). It may, however, be quite interesting to consider nonparametric error distributions, for example, mixtures of Dirichlet processes; such a work in the linear cases has been undertaken by Caron, Davy, Doucet, Duflos & Vanheeghe (2007). We shall also consider matrix-variate extensions to our current work,

in addition to nonparametric error distributions. Computations for these extensions may be burdensome in the extreme, but we hope that the novel TMCMC theory proposed by Dutta & Bhattacharya (2011) will be very useful in this regard.

Supplement to “Bayesian Inference in Nonparametric Dynamic State-Space Models”

Throughout, we refer to our main paper Ghosh, Mukhopadhyay, Roy & Bhattacharya (2011a) as GMRB.

S-9. DETAILS OF MCMC SAMPLING IN THE UNIVARIATE SITUATION

S-9.1 Updating β_f using Gibbs step

The full conditional of β_f is m -variate normal with mean

$$E[\beta_f | \dots] = \left\{ \mathbf{H}'_{D_T} (\sigma_f^2 \mathbf{A}_{f,D_T} + \sigma_\epsilon^2 \mathbf{I})^{-1} \mathbf{H}_{D_T} + \Sigma_{\beta_f,0}^{-1} \right\}^{-1} \times \left\{ \mathbf{H}'_{D_T} (\sigma_f^2 \mathbf{A}_{f,D_T} + \sigma_\epsilon^2 \mathbf{I})^{-1} \mathbf{D}_T + \Sigma_{\beta_f,0}^{-1} \beta_{f,0} \right\}. \quad (76)$$

and variance

$$V[\beta_f | \dots] = \left\{ \mathbf{H}'_{D_T} (\sigma_f^2 \mathbf{A}_{f,D_T} + \sigma_\epsilon^2 \mathbf{I})^{-1} \mathbf{H}_{D_T} + \Sigma_{\beta_f,0}^{-1} \right\}^{-1}. \quad (77)$$

S-9.2 Updating β_g using Gibbs step

The conditional distribution of β_g is m -variate normal with mean

$$\begin{aligned} E[\beta_g | \dots] &= \left\{ \Sigma_{\beta_g,0}^{-1} + \frac{\mathbf{H}'_{D_{z0}} \mathbf{A}_{g,D_{z0}}^{-1} \mathbf{H}_{D_{z0}}}{\sigma_g^2} \right. \\ &\quad \left. + \sum_{t=1}^T \frac{(\mathbf{H}'_{D_z} \mathbf{A}_{g,D_z}^{-1} \mathbf{s}_{g,D_z}(x_{t+1,t}^*) - \mathbf{h}(x_{t+1,t}^*)) (\mathbf{H}'_{D_z} \mathbf{A}_{g,D_z}^{-1} \mathbf{s}_{g,D_z}(x_{t+1,t}^*) - \mathbf{h}(x_{t+1,t}^*))'}{\sigma_{g,\eta,t}^2} \right\}^{-1} \\ &\quad \times \left\{ \Sigma_{\beta_g,0}^{-1} \beta_{g,0} + \frac{\mathbf{H}'_{D_{z0}} \mathbf{A}_{g,D_{z0}}^{-1} \mathbf{D}_{z0}}{\sigma_g^2} \right. \\ &\quad \left. + \sum_{t=1}^T \frac{(x_{t+1} - \mathbf{s}_{g,D_z}(x_{t+1,t}^*)' \mathbf{A}_{g,D_z}^{-1} \mathbf{D}_z) (\mathbf{h}(x_{t+1,t}^*) - \mathbf{H}'_{D_z} \mathbf{A}_{g,D_z}^{-1} \mathbf{s}_{g,D_z}(x_{t+1,t}^*))}{\sigma_{g,\eta,t}^2} \right\}. \quad (78) \end{aligned}$$

and variance

$$\begin{aligned}
V[\boldsymbol{\beta}_g \mid \dots] &= \left\{ \boldsymbol{\Sigma}_{\beta_g,0}^{-1} + \frac{\mathbf{H}'_{D_{z0}} \mathbf{A}_{g,D_{z0}}^{-1} \mathbf{H}_{D_{z0}}}{\sigma_g^2} \right. \\
&\quad \left. + \sum_{t=1}^T \frac{(\mathbf{H}'_{D_z} \mathbf{A}_{g,D_z}^{-1} \mathbf{s}_{g,D_z}(x_{t+1,t}^*) - \mathbf{h}(x_{t+1,t}^*)) (\mathbf{H}'_{D_z} \mathbf{A}_{g,D_z}^{-1} \mathbf{s}_{g,D_z}(x_{t+1,t}^*) - \mathbf{h}(x_{t+1,t}^*))'}{\sigma_{g,\eta,t}^2} \right\}^{-1}
\end{aligned} \tag{79}$$

In (78) and (79),

$$\sigma_{g,\eta,t}^2 = \sigma_g^2 \{1 - \mathbf{s}_{g,D_z}(x_{t+1,t}^*)' \mathbf{A}_{g,D_z}^{-1} \mathbf{s}_{g,D_z}(x_{t+1,t}^*)\} + \sigma_\eta^2. \tag{80}$$

S-9.3 Updating σ_f^2 and σ_g^2 using MH steps

The full conditionals of σ_f^2 and σ_g^2 are not available in closed forms and MH steps are necessary here. For proposal distributions we first construct a new model by setting $\sigma_\epsilon^2 = \sigma_f^2$ and $\sigma_\eta^2 = \sigma_g^2$. For this model the full conditional distributions of σ_f^2 and σ_g^2 are inverse Gamma distributions, given by

$$q_{\sigma_f^2}(\sigma_f^2) \propto (\sigma_f^2)^{-\left(\frac{T+\alpha_f+2}{2}\right)} \exp \left[-\frac{1}{2\sigma_f^2} \left\{ \gamma_f + (\mathbf{D}_T - \mathbf{H}_{D_T} \boldsymbol{\beta}_f)' (\mathbf{A}_{f,D_T} + \mathbf{I}_T)^{-1} (\mathbf{D}_T - \mathbf{H}_{D_T} \boldsymbol{\beta}_f) \right\} \right] \tag{81}$$

and

$$\begin{aligned}
q_{\sigma_g^2}(\sigma_g^2) &\propto (\sigma_g^2)^{-\left(\frac{\alpha_g+4+n+T}{2}\right)} \exp \left[-\frac{1}{2\sigma_g^2} \left\{ \gamma_g + (x_1 - g(x_{1,0}^*))^2 \right. \right. \\
&\quad \left. + \sum_{t=1}^T \frac{\{x_{t+1} - \mathbf{h}(x_{t+1,t}^*)' \boldsymbol{\beta}_g - \mathbf{s}_{g,D_z}(x_{t+1,t}^*)' \mathbf{A}_{g,D_z}^{-1} (\mathbf{D}_z - \mathbf{H}_{D_z} \boldsymbol{\beta}_g)\}^2}{\sigma_{g,t}^2} \right. \\
&\quad \left. \left. + (\mathbf{D}_{z0} - \mathbf{H}_{D_{z0}} \boldsymbol{\beta}_g)' \mathbf{A}_{g,D_{z0}}^{-1} (\mathbf{D}_{z0} - \mathbf{H}_{D_{z0}} \boldsymbol{\beta}_g) \right\} \right]
\end{aligned} \tag{82}$$

In (82),

$$\sigma_{g,t}^2 = 2 - \mathbf{s}_{g,D_z}(x_{t+1,t}^*)' \mathbf{A}_{g,D_z}^{-1} \mathbf{s}_{g,D_z}(x_{t+1,t}^*). \tag{83}$$

In (82) the term $(\sigma_g^2)^{-1/2} \exp \left\{ -\frac{1}{2\sigma_g^2} (x_1 - g(x_{1,0}^*))^2 \right\}$ is also taken into account since, given $\sigma_\eta^2 = \sigma_g^2$, $[x_1 | g(x_{1,0}^*), x_0] \sim N(g(x_{1,0}^*), \sigma_g^2)$. It is useful to remark here that unless $\sigma_f \approx \sigma_\epsilon$ and $\sigma_g \approx \sigma_\eta$ these proposal mechanisms may not be efficient. We shall discuss other proposal distributions in the context of applications.

S-9.4 Updating σ_ϵ^2 and σ_η^2 using MH steps

As before, the full conditionals of σ_ϵ^2 and σ_η^2 are not available in closed forms. For MH steps, we construct proposal distributions obtained by setting $\sigma_f^2 = \sigma_\epsilon^2$ and $\sigma_g^2 = \sigma_\eta^2$. The proposal distributions are given by

$$q_{\sigma_\epsilon^2}(\sigma_\epsilon^2) \propto (\sigma_\epsilon^2)^{-\left(\frac{T+\alpha_\epsilon+2}{2}\right)} \exp \left[-\frac{1}{2\sigma_\epsilon^2} \left\{ \gamma_\epsilon + (\mathbf{D}_T - \mathbf{H}_{D_T} \boldsymbol{\beta}_f)' (\mathbf{A}_{f,D_T} + \mathbf{I}_T)^{-1} (\mathbf{D}_T - \mathbf{H}_{D_T} \boldsymbol{\beta}_f) \right\} \right] \quad (84)$$

$$\begin{aligned} q_{\sigma_\eta^2}(\sigma_\eta^2) \propto (\sigma_\eta^2)^{-\left(\frac{\alpha_\eta+4+n+T}{2}\right)} \exp \left[-\frac{1}{2\sigma_\eta^2} \left\{ \gamma_\eta + (x_1 - g(x_{1,0}^*))^2 \right. \right. \\ \left. \left. + (\mathbf{D}_{z0} - \mathbf{H}_{D_{z0}} \boldsymbol{\beta}_g)' \mathbf{A}_{g,D_{z0}}^{-1} (\mathbf{D}_{z0} - \mathbf{H}_{D_{z0}} \boldsymbol{\beta}_g) \right. \right. \\ \left. \left. + \sum_{t=1}^T \frac{\left\{ x_{t+1} - \mathbf{h}(x_{t+1,t}^*)' \boldsymbol{\beta}_g - \mathbf{s}_{g,D_z}(x_{t+1,t}^*)' \mathbf{A}_{g,D_z}^{-1} (\mathbf{D}_z - \mathbf{H}_{D_z} \boldsymbol{\beta}_g) \right\}^2}{\sigma_{g,t}^2} \right\} \right] \end{aligned} \quad (85)$$

In (85) the term $(\sigma_\eta^2)^{-(n+1)/2} \exp \left\{ -\frac{1}{2\sigma_\eta^2} (\mathbf{D}_{z0} - \mathbf{H}_{D_{z0}} \boldsymbol{\beta}_g)' \mathbf{A}_{g,D_{z0}}^{-1} (\mathbf{D}_{z0} - \mathbf{H}_{D_{z0}} \boldsymbol{\beta}_g) \right\}$ occurs since, given $\sigma_g^2 = \sigma_\eta^2$, $[\mathbf{D}_{z0} | \boldsymbol{\beta}_g, \sigma_g^2, \mathbf{R}_g] \sim N_{n+1}(\mathbf{H}_{D_{z0}} \boldsymbol{\beta}_g, \sigma_\eta^2 \mathbf{A}_{g,D_{z0}})$. Again, these proposals need not be efficient unless $\sigma_f \approx \sigma_\epsilon$ and $\sigma_g \approx \sigma_\eta$. In the context of specific applications we shall discuss other proposal distributions.

S-9.5 Updating \mathbf{R}_f and \mathbf{R}_g using MH steps

For $i = 1, 2$, the full conditionals of the smoothness parameters $r_{i,f}$ and $r_{i,g}$ are not available in closed forms, and we suggest MH steps with normal random walk proposals with adequately optimized variances.

S-9.6 Updating $g(x_{1,0}^*)$ using Gibbs step

The full conditional of $g(x_{1,0}^*)$ is univariate normal with mean and variance given, respectively, by

$$E[g(x_{1,0}^*) \mid \cdots] = \left\{ \frac{1}{\sigma_\eta^2} + \frac{1 + \mathbf{s}_{g,D_z}(x_{1,0}^*)' \boldsymbol{\Sigma}_{g,D_z}^{-1} \mathbf{s}_{g,D_z}(x_{1,0}^*)}{\sigma_g^2} \right\}^{-1} \\ \times \left\{ \frac{x_1}{\sigma_\eta^2} + \frac{\mathbf{h}(x_{1,0}^*)' \boldsymbol{\beta}_g + \mathbf{s}_{g,D_z}(x_{1,0}^*)' \boldsymbol{\Sigma}_{g,D_z}^{-1} \mathbf{D}_z^*}{\sigma_g^2} \right\} \quad (86)$$

and

$$V[g(x_{1,0}^*) \mid \cdots] = \left\{ \frac{1}{\sigma_\eta^2} + \frac{1 + \mathbf{s}_{g,D_z}(x_{1,0}^*)' \boldsymbol{\Sigma}_{g,D_z}^{-1} \mathbf{s}_{g,D_z}(x_{1,0}^*)}{\sigma_g^2} \right\}^{-1} \quad (87)$$

In (86),

$$\mathbf{D}_z^* = \mathbf{D}_z - \mathbf{H}_{D_z} \boldsymbol{\beta}_g + \mathbf{s}_{g,D_z}(x_{1,0}^*) \mathbf{h}(x_{1,0}^*)' \boldsymbol{\beta}_g \quad (88)$$

S-9.7 Updating \mathbf{D}_z using Gibbs step

The full conditional distribution of \mathbf{D}_z is n -variate normal with mean

$$E[\mathbf{D}_z \mid \cdots] = \left\{ \frac{\boldsymbol{\Sigma}_{g,D_z}^{-1}}{\sigma_g^2} + \mathbf{A}_{g,D_z}^{-1} \left(\sum_{t=1}^T \frac{\mathbf{s}_{g,D_z}(x_{t+1,t}^*) \mathbf{s}_{g,D_z}(x_{t+1,t}^*)'}{\sigma_{g,\eta,t}^2} \right) \mathbf{A}_{g,D_z}^{-1} \right\}^{-1} \\ \times \left\{ \frac{\boldsymbol{\Sigma}_{g,D_z}^{-1} \boldsymbol{\mu}_{g,D_z}}{\sigma_g^2} + \mathbf{A}_{g,D_z}^{-1} \sum_{t=1}^T \frac{\mathbf{s}_{g,D_z}(x_{t+1,t}^*) \{x_{t+1} - \boldsymbol{\beta}_g' (\mathbf{h}(x_{t+1,t}^*) - \mathbf{H}_{D_z}' \mathbf{A}_{g,D_z}^{-1} \mathbf{s}_{g,D_z}(x_{t+1,t}^*))\}}{\sigma_{g,\eta,t}^2} \right\} \quad (89)$$

and variance

$$V[\mathbf{D}_z \mid \cdots] = \left\{ \frac{\boldsymbol{\Sigma}_{g,D_z}^{-1}}{\sigma_g^2} + \mathbf{A}_{g,D_z}^{-1} \left(\sum_{t=1}^T \frac{\mathbf{s}_{g,D_z}(x_{t+1,t}^*) \mathbf{s}_{g,D_z}(x_{t+1,t}^*)'}{\sigma_{g,\eta,t}^2} \right) \mathbf{A}_{g,D_z}^{-1} \right\}^{-1} \quad (90)$$

In (89) and (90), $\boldsymbol{\mu}_{g,D_z}$ and $\boldsymbol{\Sigma}_{g,D_z}$ are given by (9) and (10), respectively, of GMRB.

S-9.8 Updating x_0 using MH step

Let $\beta_f = (\beta'_{0,f}, \beta_{1,f})'$ and $\beta_g = (\beta'_{0,g}, \beta_{1,g})'$, where $\beta_{0,f}$ and $\beta_{0,g}$ are two-component vectors. Assuming the prior of x_0 to be normal with mean μ_{x_0} and variance $\sigma_{x_0}^2$, and setting $\sigma_g^2 = 0$, the full conditional distribution of x_1 is univariate normal, with mean and variance given, respectively, by

$$\xi_0 = \left(\frac{1}{\sigma_{x_0}^2} + \frac{\beta_{1,g}^2}{\sigma_\eta^2} \right)^{-1} \left\{ \frac{\mu_{x_0}}{\sigma_{x_0}^2} + \frac{(x_1 - \mathbf{k}'_1 \beta_{0,g}) \beta_{1,g}}{\sigma_\eta^2} \right\}, \quad (91)$$

$$\phi_0^2 = \left(\frac{1}{\sigma_{x_0}^2} + \frac{\beta_{1,g}^2}{\sigma_\eta^2} \right)^{-1}. \quad (92)$$

In (91), for any t , $\mathbf{k}_t = (1, t)'$. We use $q_{x_0}(x_0) \equiv N(x_0 : \xi_0, \phi_0^2)$ as the proposal distribution for updating x_0 using MH step. Observe that, under $\sigma_g^2 = 0$, $g(x_{1,0}^*) = \mathbf{h}(x_{1,0}^*)' \beta_g$ with probability one; hence $[x_1 | g(x_{1,0}^*), x_0, \beta_g, \sigma_\eta^2] \sim N(\mathbf{h}(x_{1,0}^*)' \beta_g, \sigma_\eta^2)$, which has been taken into account while constructing the above proposal distribution. Note that this proposal will only be efficient when the $g(\cdot, \cdot)$ is close to linear. As a result, for non-linear applications, we shall often use other proposal mechanisms, such as the normal random walk.

S-9.9 Updating $\{x_1, \dots, x_T\}$ using MH steps

We construct proposal distributions for simulating $\{x_1, \dots, x_T\}$ based on linear observational and evolutionary equations, setting $\sigma_f^2 = \sigma_g^2 = 0$. Thus, for $t = 0, \dots, T-1$, the proposal distributions of x_{t+1} are of the form $q_{x_{t+1}}(x_{t+1}) \equiv N(x_{t+1} : \xi_{t+1}, \phi_{t+1}^2)$, that is, a normal distribution with mean ξ_{t+1} and variance ϕ_{t+1}^2 , where the latter quantities are given by

$$\xi_{t+1} = \left(\frac{1 + \beta_{1,g}^2}{\sigma_\eta^2} + \frac{\beta_{1,f}^2}{\sigma_\epsilon^2} \right)^{-1} \left\{ \frac{\mathbf{k}'_{t+1} \beta_{0,g} + \beta_{1,g} x_t + (x_{t+2} - \mathbf{k}'_{t+2} \beta_{0,g}) \beta_{1,g}}{\sigma_\eta^2} + \frac{(y_{t+1} - \mathbf{k}'_{t+1} \beta_{0,f}) \beta_{1,f}}{\sigma_\epsilon^2} \right\} \quad (93)$$

$$\phi_{t+1}^2 = \left(\frac{1 + \beta_{1,g}^2}{\sigma_\eta^2} + \frac{\beta_{1,f}^2}{\sigma_\epsilon^2} \right)^{-1} \quad (94)$$

These proposal mechanisms will be efficient only if both $f(\cdot, \cdot)$ and $g(\cdot, \cdot)$ are close to linear. Hence, for non-linear situations, we shall consider other proposal distributions.

S-9.10 Updating x_{T+1} using Gibbs step

The full conditional distribution of x_{T+1} is normal with mean and variance given, respectively, by

$$E[x_{T+1} \mid \cdots] = \mathbf{h}(x_{T+1,T}^*)' \boldsymbol{\beta}_g + \mathbf{s}_{g,D_z}(x_{T+1,T}^*)' \mathbf{A}_{g,D_z}^{-1} (\mathbf{D}_z - \mathbf{H}_{D_z} \boldsymbol{\beta}_g) \quad (95)$$

and variance

$$V[x_{T+1} \mid \cdots] = \sigma_\eta^2 + \sigma_g^2 \{1 - \mathbf{s}_{g,D_z}(x_{T+1,T}^*)' \mathbf{A}_{g,D_z}^{-1} \mathbf{s}_{g,D_z}(x_{T+1,T}^*)\} \quad (96)$$

S-10. DETAILS OF MCMC SAMPLING IN THE MULIVARIATE SITUATION

To obtain good proposal distributions, we use our old strategy of ignoring the error terms $\boldsymbol{\epsilon}_t$ and $\boldsymbol{\eta}_t$. Below we provide details of the proposal distributions used in each case. For our purpose, we abuse notation slightly as in Section 4 of GMRB, that is, we denote by \mathbf{G}_{z0} , \mathbf{D}_{z0} , $\mathbf{A}_{g,z0}$ and $\mathbf{H}_{g,D_{z0}}$ the multivariate analogues of the quantities corresponding to the univariate situation described in Section 4 of GMRB. In other words, let $\mathbf{G}_{z0} = \mathbf{G} \cup \{\mathbf{x}_{1,0}^*\}$, $\mathbf{D}_{z0} = (\mathbf{D}_z', \mathbf{g}(\mathbf{x}_{1,0}^*))'$, $\mathbf{A}_{g,z0} =$

$$\begin{pmatrix} \mathbf{A}_{g,D_z} & \mathbf{s}_{g,D_z}(\mathbf{x}_{1,0}^*) \\ \mathbf{s}_{g,D_z}(\mathbf{x}_{1,0}^*)' & 1 \end{pmatrix} \text{ and } \mathbf{H}_{g,D_{z0}}' = [\mathbf{H}_{g,D_z}', \mathbf{h}(\mathbf{x}_{1,0}^*)].$$

S-10.1 Proposal distribution for updating \mathbf{B}_f

Assuming $\boldsymbol{\Sigma}_\epsilon = \boldsymbol{\Sigma}_f$ in our multivariate dynamic state-space model, the full conditional of \mathbf{B}_f is $m \times p$ -variate matrix-normal:

$$\mathbf{B}_f \sim \mathcal{N}_{m,p}(\boldsymbol{\mu}_{B_f}, \boldsymbol{\Sigma}_{B_f}, \boldsymbol{\Sigma}_f), \quad (97)$$

where

$$\begin{aligned} \boldsymbol{\mu}_{B_f} &= \left(\mathbf{H}_{D_T}' (\mathbf{A}_{f,D_T} + \mathbf{I}_T)^{-1} \mathbf{H}_{D_T} + \psi^{-1} \boldsymbol{\Sigma}_{B_f,0}^{-1} \right)^{-1} \\ &\quad \times \left(\mathbf{H}_{D_T}' (\mathbf{A}_{f,D_T} + \mathbf{I}_T)^{-1} \mathbf{D}_T + \psi^{-1} \boldsymbol{\Sigma}_{B_f,0}^{-1} \mathbf{B}_{f,0} \right), \end{aligned} \quad (98)$$

and

$$\boldsymbol{\Sigma}_{B_f} = \left(\mathbf{H}_{D_T}' (\mathbf{A}_{f,D_T} + \mathbf{I}_T)^{-1} \mathbf{H}_{D_T} + \psi^{-1} \boldsymbol{\Sigma}_{B_f,0}^{-1} \right)^{-1}. \quad (99)$$

S-10.2 Proposal distribution for \mathbf{B}_g

Assuming $\Sigma_\eta = \Sigma_g$, the conditional distribution of \mathbf{B}_g is $m \times q$ -variate matrix:

$$\mathbf{B}_g \sim \mathcal{N}_{m,q} \left(\boldsymbol{\mu}_{B_g}, \Sigma_{B_g}, \Sigma_g \right), \quad (100)$$

where

$$\begin{aligned} \boldsymbol{\mu}_{B_g} = & \left\{ \psi^{-1} \Sigma_{B_g,0}^{-1} + \mathbf{H}'_{D_{z0}} \mathbf{A}_{g,D_{z0}}^{-1} \mathbf{H}_{D_{z0}} \right. \\ & + \sum_{t=1}^T \frac{(\mathbf{H}'_{D_z} \mathbf{A}_{g,D_z}^{-1} \mathbf{s}_{g,D_z}(\mathbf{x}_{t+1,t}^*) - \mathbf{h}(\mathbf{x}_{t+1,t}^*)) (\mathbf{H}'_{D_z} \mathbf{A}_{g,D_z}^{-1} \mathbf{s}_{g,D_z}(\mathbf{x}_{t+1,t}^*) - \mathbf{h}(\mathbf{x}_{t+1,t}^*))'}{\sigma_{g,t}^2} \Big\}^{-1} \\ & \times \left\{ \psi^{-1} \Sigma_{B_g,0}^{-1} \mathbf{B}_{g,0} + \mathbf{H}'_{D_{z0}} \mathbf{A}_{g,D_{z0}}^{-1} \mathbf{D}_{z0} \right. \\ & + \sum_{t=1}^T \frac{(\mathbf{h}(\mathbf{x}_{t+1,t}^*) - \mathbf{H}'_{D_z} \mathbf{A}_{g,D_z}^{-1} \mathbf{s}_{g,D_z}(\mathbf{x}_{t+1,t}^*)) (\mathbf{x}_{t+1} - \mathbf{D}'_z \mathbf{A}_{g,D_z}^{-1} \mathbf{s}_{g,D_z}(\mathbf{x}_{t+1,t}^*))'}{\sigma_{g,t}^2} \Big\}. \end{aligned} \quad (101)$$

and

$$\begin{aligned} \Sigma_{B_g} = & \left\{ \psi^{-1} \Sigma_{B_g,0}^{-1} + \mathbf{H}'_{D_{z0}} \mathbf{A}_{g,D_{z0}}^{-1} \mathbf{H}_{D_{z0}} \right. \\ & + \sum_{t=1}^T \frac{(\mathbf{H}'_{D_z} \mathbf{A}_{g,D_z}^{-1} \mathbf{s}_{g,D_z}(\mathbf{x}_{t+1,t}^*) - \mathbf{h}(\mathbf{x}_{t+1,t}^*)) (\mathbf{H}'_{D_z} \mathbf{A}_{g,D_z}^{-1} \mathbf{s}_{g,D_z}(\mathbf{x}_{t+1,t}^*) - \mathbf{h}(\mathbf{x}_{t+1,t}^*))'}{\sigma_{g,t}^2} \Big\}^{-1} \end{aligned} \quad (102)$$

In (101) and (102), $\sigma_{g,t}^2$ is given by (83), with obvious notational change from the univariate x_t to the multivariate \mathbf{x}_t . The corresponding changes from x^* to \mathbf{x}^* are also self-explanatory.

S-10.3 Proposal distributions for updating Σ_f and Σ_g

Setting $\Sigma_\epsilon = \Sigma_f$ and $\Sigma_\eta = \Sigma_g$, we obtain the following inverse Wishart proposal distributions of Σ_f and Σ_g :

$$\begin{aligned} q_{\Sigma_f}(\Sigma_f) \propto & |\Sigma_f|^{-\left(\frac{\nu_f + p + 1 + T + m}{2}\right)} \exp \left[-\frac{1}{2} \text{tr} \left\{ \Sigma_f^{-1} \left(\Sigma_{f,0} + \psi^{-1} (\mathbf{B}_f - \mathbf{B}_{f,0})' \Sigma_{B_f,0}^{-1} (\mathbf{B}_f - \mathbf{B}_{f,0}) \right. \right. \right. \\ & \left. \left. \left. + (\mathbf{D}_T - \mathbf{H}_{D_T} \mathbf{B}_f)' (\mathbf{A}_{f,D_T} + \mathbf{I}_T)^{-1} (\mathbf{D}_T - \mathbf{H}_{D_T} \mathbf{B}_f) \right) \right\} \right] \end{aligned} \quad (103)$$

and

$$\begin{aligned}
q_{\Sigma_g}(\Sigma_g) \propto |\Sigma_g|^{-\left(\frac{\nu_g+q+2+m+n+T}{2}\right)} \exp \left[-\frac{1}{2} \text{tr} \Sigma_g^{-1} \left\{ \Sigma_{g,0} + \psi^{-1} (\mathbf{B}_g - \mathbf{B}_{g,0})' \Sigma_{B_g,0}^{-1} (\mathbf{B}_g - \mathbf{B}_{g,0}) \right. \right. \\
+ (\mathbf{x}_1 - \mathbf{g}(\mathbf{x}_{1,0}^*)) (\mathbf{x}_1 - \mathbf{g}(\mathbf{x}_{1,0}^*))' \\
+ \sum_{t=1}^T \frac{1}{\sigma_{g,t}^2} \left\{ \mathbf{x}_{t+1} - \mathbf{B}_g' \mathbf{h}(\mathbf{x}_{t+1,t}^*) - (\mathbf{D}_z - \mathbf{H}_{D_z} \mathbf{B}_g)' \mathbf{A}_{g,D_z}^{-1} \mathbf{s}_{g,D_z}(\mathbf{x}_{t+1,t}^*) \right\} \\
\times \left\{ \mathbf{x}_{t+1} - \mathbf{B}_g' \mathbf{h}(\mathbf{x}_{t+1,t}^*) - (\mathbf{D}_z - \mathbf{H}_{D_z} \mathbf{B}_g)' \mathbf{A}_{g,D_z}^{-1} \mathbf{s}_{g,D_z}(\mathbf{x}_{t+1,t}^*) \right\}' \\
\left. \left. + (\mathbf{D}_{z0} - \mathbf{H}_{D_{z0}} \mathbf{B}_g)' \mathbf{A}_{g,D_{z0}}^{-1} (\mathbf{D}_{z0} - \mathbf{H}_{D_{z0}} \mathbf{B}_g) \right\} \right] \quad (104)
\end{aligned}$$

S-10.4 Proposal distributions for updating Σ_ϵ , Σ_η , \mathbf{R}_f and \mathbf{R}_g

As in the univariate case, here we set $\Sigma_f = \Sigma_\epsilon$ and $\Sigma_g = \Sigma_\eta$. Then the proposal distributions of Σ_ϵ and Σ_η are given by the following:

$$\begin{aligned}
q_{\Sigma_\epsilon}(\Sigma_\epsilon) \propto |\Sigma_\epsilon|^{-\left(\frac{\nu_\epsilon+p+1+T}{2}\right)} \\
\times \exp \left[-\frac{1}{2} \text{tr} \Sigma_\epsilon^{-1} \left\{ \Sigma_{\epsilon,0} + (\mathbf{D}_T - \mathbf{H}_{D_T} \mathbf{B}_f)' (\mathbf{A}_{f,D_T} + \mathbf{I}_T)^{-1} (\mathbf{D}_T - \mathbf{H}_{D_T} \mathbf{B}_f) \right\} \right] \quad (105)
\end{aligned}$$

$$\begin{aligned}
q_{\Sigma_\eta}(\Sigma_\eta) \propto |\Sigma_\eta|^{-\left(\frac{\nu_\eta+q+3+n+T}{2}\right)} \exp \left[-\frac{1}{2} \text{tr} \Sigma_\eta^{-1} \left\{ \Sigma_{\eta,0} + (\mathbf{x}_1 - \mathbf{g}(\mathbf{x}_{1,0}^*)) (\mathbf{x}_1 - \mathbf{h}(\mathbf{x}_{1,0}^*))' \right. \right. \\
+ (\mathbf{D}_{z0} - \mathbf{H}_{D_{z0}} \mathbf{B}_g)' \mathbf{A}_{g,D_{z0}}^{-1} (\mathbf{D}_{z0} - \mathbf{H}_{D_{z0}} \mathbf{B}_g) \\
+ \sum_{t=1}^T \frac{1}{\sigma_{g,t}^2} \left\{ \mathbf{x}_{t+1} - \mathbf{B}_g' \mathbf{h}(\mathbf{x}_{t+1,t}^*) - (\mathbf{D}_z - \mathbf{H}_{D_z} \mathbf{B}_g)' \mathbf{A}_{g,D_z}^{-1} \mathbf{s}_{g,D_z}(\mathbf{x}_{t+1,t}^*) \right\} \\
\left. \left. \times \left\{ \mathbf{x}_{t+1} - \mathbf{B}_g' \mathbf{h}(\mathbf{x}_{t+1,t}^*) - (\mathbf{D}_z - \mathbf{H}_{D_z} \mathbf{B}_g)' \mathbf{A}_{g,D_z}^{-1} \mathbf{s}_{g,D_z}(\mathbf{x}_{t+1,t}^*) \right\}' \right\} \right] \quad (106)
\end{aligned}$$

As in the case of the corresponding univariate proposals, in (106) the factor

$$|\Sigma_\eta|^{-\left(\frac{n}{2}\right)} \exp \left[-\frac{1}{2} \text{tr} \Sigma_\eta^{-1} (\mathbf{D}_{z0} - \mathbf{H}_{D_{z0}} \mathbf{B}_g)' \mathbf{A}_{g,D_{z0}}^{-1} (\mathbf{D}_{z0} - \mathbf{H}_{D_{z0}} \mathbf{B}_g) \right]$$

occurs because under $\Sigma_g = \Sigma_\eta$, $[\mathbf{D}_{z0} \mid \mathbf{B}_g, \Sigma_g, \mathbf{R}_g] \sim \mathcal{N}_{n,q}(\mathbf{H}_{D_{z0}} \mathbf{B}_g, \mathbf{A}_{g,D_{z0}}, \Sigma_\eta)$.

The full conditionals of the smoothness parameters in \mathbf{R}_f and \mathbf{R}_g are not available in closed

forms, and as before we suggest Metropolis-Hastings steps with normal random walk proposals with adequately optimized variances.

S-10.5 Proposal distribution for updating $\mathbf{g}(\mathbf{x}_{1,0}^*)$

Assuming $\Sigma_\eta = \Sigma_g$, the full conditional of $\mathbf{g}(\mathbf{x}_{1,0}^*)$ is q -variate normal with mean and variance given, respectively, by

$$\begin{aligned} \boldsymbol{\mu}_g^* &= E[\mathbf{g}(\mathbf{x}_{1,0}^*) \mid \cdots] = \{2 + \mathbf{s}_{g,D_z}(\mathbf{x}_{1,0}^*)' \Sigma_{g,D_z}^{-1} \mathbf{s}_{g,D_z}(\mathbf{x}_{1,0}^*)\}^{-1} \Sigma_g \\ &\quad \times \{\mathbf{x}_1 + \mathbf{B}_g' \mathbf{h}(\mathbf{x}_{1,0}^*) + \mathbf{D}^{*'} \Sigma_{g,D_z}^{-1} \mathbf{s}_{g,D_z}(\mathbf{x}_{1,0}^*)\} \end{aligned} \quad (107)$$

and

$$\Sigma_g^* = V[\mathbf{g}(\mathbf{x}_{1,0}^*) \mid \cdots] = \{2 + \mathbf{s}_{g,D_z}(\mathbf{x}_{1,0}^*)' \Sigma_{g,D_z}^{-1} \mathbf{s}_{g,D_z}(\mathbf{x}_{1,0}^*)\}^{-1} \Sigma_g \quad (108)$$

In the above, Σ_{g,D_z} is given by (59) of GMRB, and \mathbf{D}_z^* is given by

$$\mathbf{D}_z^* = \mathbf{D}_z - \mathbf{H}_{D_z} \mathbf{B}_g + \mathbf{s}_{g,D_z}(\mathbf{x}_{1,0}^*) \mathbf{h}(\mathbf{x}_{1,0}^*)' \mathbf{B}_g \quad (109)$$

We consider $q_{g(\mathbf{x}_{1,0}^*)}(\mathbf{g}(\mathbf{x}_{1,0}^*)) \equiv N_q(\mathbf{g}(\mathbf{x}_{1,0}^*) : \boldsymbol{\mu}_g^*, \Sigma_g^*)$ as the proposal distribution for updating $\mathbf{g}(\mathbf{x}_{1,0}^*)$.

S-10.6 Proposal distribution for updating \mathbf{D}_z

The full conditional distribution of \mathbf{D}_z in our dynamic model after setting $\Sigma_\eta = \Sigma_g$, is matrix-normal:

$$[\mathbf{D}_z] \sim \mathcal{N}_{n,q}(\boldsymbol{\mu}_{D_z}^*, \Sigma_{D_z}^*, \Sigma_g), \quad (110)$$

where

$$\begin{aligned} \boldsymbol{\mu}_{D_z}^* &= \left\{ \Sigma_{g,D_z}^{-1} + \mathbf{A}_{g,D_z}^{-1} \left(\sum_{t=1}^T \frac{\mathbf{s}_{g,D_z}(\mathbf{x}_{t+1,t}^*) \mathbf{s}_{g,D_z}(\mathbf{x}_{t+1,t}^*)'}{\sigma_{g,t}^2} \right) \mathbf{A}_{g,D_z}^{-1} \right\}^{-1} \\ &\quad \times \left\{ \Sigma_{g,D_z}^{-1} \boldsymbol{\mu}_{g,D_z} + \mathbf{A}_{g,D_z}^{-1} \sum_{t=1}^T \frac{\mathbf{s}_{g,D_z}(\mathbf{x}_{t+1,t}^*) \{ \mathbf{x}_{t+1}' - (\mathbf{h}(\mathbf{x}_{t+1,t}^*)' - \mathbf{s}_{g,D_z}(\mathbf{x}_{t+1,t}^*)' \mathbf{A}_{g,D_z}^{-1} \mathbf{H}_{D_z}) \mathbf{B}_g \}}{\sigma_{g,t}^2} \right\} \end{aligned} \quad (111)$$

and

$$\Sigma_{D_z}^* = \left\{ \Sigma_{g,D_z}^{-1} + \mathbf{A}_{g,D_z}^{-1} \left(\sum_{t=1}^T \frac{\mathbf{s}_{g,D_z}(\mathbf{x}_{t+1,t}^*) \mathbf{s}_{g,D_z}(\mathbf{x}_{t+1,t}^*)'}{\sigma_{g,t}^2} \right) \mathbf{A}_{g,D_z}^{-1} \right\}^{-1} \quad (112)$$

In the above, $\boldsymbol{\mu}_{g,D_z}$ and Σ_{g,D_z} are given by (58) and (59) of GMRB, respectively. The distribution (110) will be used as proposal distribution to update D_z using Metropolis-Hastings step.

S-10.7 MH proposal for updating \mathbf{x}_0

For $j = 1, \dots, p$, let $\boldsymbol{\beta}_{j,f} = (\boldsymbol{\beta}'_{j,f,1}, \boldsymbol{\beta}'_{j,f,2})'$. Likewise, for $j = 1, \dots, q$, let $\boldsymbol{\beta}_{j,g} = (\boldsymbol{\beta}'_{j,g,1}, \boldsymbol{\beta}'_{j,g,2})'$. Here $\boldsymbol{\beta}_{j,f,1}$ and $\boldsymbol{\beta}_{j,g,1}$ are two-dimensional and $\boldsymbol{\beta}_{j,f,2}$ and $\boldsymbol{\beta}_{j,g,2}$ are q -dimensional vectors. Also let $\mathbf{a}_f = (\boldsymbol{\beta}_{1,f,1}, \dots, \boldsymbol{\beta}_{p,f,1})'$ and $\mathbf{b}_f = (\boldsymbol{\beta}_{1,f,2}, \dots, \boldsymbol{\beta}_{p,f,2})'$. Similarly, $\mathbf{a}_g = (\boldsymbol{\beta}_{1,g,1}, \dots, \boldsymbol{\beta}_{q,g,1})'$ and $\mathbf{b}_g = (\boldsymbol{\beta}_{1,g,2}, \dots, \boldsymbol{\beta}_{q,g,2})'$. Thus, \mathbf{a}_f and \mathbf{a}_g are $p \times 2$ and $q \times 2$ -dimensional matrices respectively, while \mathbf{b}_f and \mathbf{b}_g are, respectively, $p \times q$ and $q \times q$ dimensional matrices. Then, $\mathbf{B}'_f \mathbf{h}(\mathbf{x}^*) = \mathbf{a}_f \mathbf{k}_1 + \mathbf{b}_f \mathbf{x}$ and $\mathbf{B}'_g \mathbf{h}(\mathbf{x}^*) = \mathbf{a}_g \mathbf{k}_1 + \mathbf{b}_g \mathbf{x}$, for any q -dimensional \mathbf{x} . With these representations, and setting $\Sigma_g = 0$, the full conditional distribution of \mathbf{x}_0 is q -variate normal with mean and variance given, respectively, by

$$E[\mathbf{x}_0 | \dots] = \{ \Sigma_{x_0}^{-1} + \mathbf{b}'_g \Sigma_{\eta}^{-1} \mathbf{b}_g \}^{-1} \{ \Sigma_{x_0}^{-1} \boldsymbol{\mu}_{x_0} + \mathbf{b}'_g \Sigma_{\eta}^{-1} (\mathbf{x}_1 - \mathbf{a}_g \mathbf{k}_1) \} \quad (113)$$

$$V[\mathbf{x}_0 | \dots] = \{ \Sigma_{x_0}^{-1} + \mathbf{b}'_g \Sigma_{\eta}^{-1} \mathbf{b}_g \}^{-1}. \quad (114)$$

S-10.8 MH proposals for $\{\mathbf{x}_1, \dots, \mathbf{x}_T\}$

Ignoring the errors of the functions \mathbf{f} and \mathbf{g} , that is, setting $\Sigma_f = \Sigma_g = 0$, it turns out that the proposal distribution of \mathbf{x}_{t+1} , for $t = 0, \dots, T-1$, can be taken as a q -variate normal distribution with mean and variance given, respectively, by

$$\begin{aligned} E[\mathbf{x}_{t+1} | \dots] &= (\Sigma_{\eta}^{-1} + \mathbf{b}'_g \Sigma_{\eta}^{-1} \mathbf{b}'_g + \mathbf{b}'_f \Sigma_{\epsilon}^{-1} \mathbf{b}_f)^{-1} \\ &\quad \times \{ \Sigma_{\eta}^{-1} (\mathbf{a}_g \mathbf{k}_{t+1} + \mathbf{b}_g \mathbf{x}_t) + \mathbf{b}'_g \Sigma_{\eta}^{-1} (\mathbf{x}_{t+2} - \mathbf{a}_g \mathbf{k}_{t+2}) + \mathbf{b}'_f \Sigma_{\epsilon}^{-1} (\mathbf{y}_{t+1} - \mathbf{a}_f \mathbf{k}_{t+1}) \} \end{aligned} \quad (115)$$

$$V[\mathbf{x}_{t+1} | \dots] = (\Sigma_{\eta}^{-1} + \mathbf{b}'_g \Sigma_{\eta}^{-1} \mathbf{b}'_g + \mathbf{b}'_f \Sigma_{\epsilon}^{-1} \mathbf{b}_f)^{-1}. \quad (116)$$

S-10.9 Gibbs step for updating \mathbf{x}_{T+1}

The full conditional distribution of \mathbf{x}_{T+1} is q -variate normal with mean

$$E[\mathbf{x}_{T+1} \mid \cdots] = \mathbf{B}'_g \mathbf{h}(\mathbf{x}_{T+1,T}^*) + (\mathbf{D}_z - \mathbf{H}_{D_z} \mathbf{B}_g)' \mathbf{A}_{g,D_z}^{-1} \mathbf{s}_{g,D_z}(\mathbf{x}_{T+1,T}^*) \quad (117)$$

and variance

$$V[\mathbf{x}_{T+1} \mid \cdots] = \Sigma_\eta + \Sigma_g \{1 - \mathbf{s}_{g,D_z}(\mathbf{x}_{T+1,T}^*)' \mathbf{A}_{g,D_z}^{-1} \mathbf{s}_{g,D_z}(\mathbf{x}_{T+1,T}^*)\}. \quad (118)$$

The caveat for using these proposal distributions is that if the underlying assumptions $\Sigma_f = \Sigma_\epsilon$, $\Sigma_g = \Sigma_\eta$ and $\Sigma_g = \mathbf{0}$ do not hold, then the proposal mechanisms will turn out to be much less effective, and more so compared to the univariate cases, due to the curse of dimensionality. So, we shall also consider other approaches to these updating procedures, to be discussed in the context of the specific applications.

REFERENCES

- Bhattacharya, S. (2007), “A Simulation Approach to Bayesian Emulation of Complex Dynamic Computer Models,” *Bayesian Analysis*, 2, 783–816.
- Brockwell, P. J., & Davis, R. A. (1987), *Time Series: Theory and Methods*, New York: Springer-Verlag.
- Carlin, B. P., Polson, N. G., & Stoffer, D. S. (1992), “A Monte Carlo Approach to Nonnormal and Nonlinear State-Space Modeling,” *Journal of the American Statistical Association*, 87, 493–500.
- Caron, F., Davy, M., Doucet, A., Duflos, E., & Vanheeghe, P. (2007), “Bayesian Inference for Linear Dynamic Models with Dirichlet Process Mixtures,”. Available at arXiv:math/0702225v1.
- Carvalho, C. M., & West, M. (2007), “Dynamic Matrix-Variate Graphical Models,” *Bayesian Analysis*, 2, 69–98.
- Cressie, N. A. C. (1993), *Statistics for Spatial Data*, New York: Wiley.
- Durbin, J., & Koopman, S. J. (2001), *Time Series Analysis by State Space Methods*, Oxford: Oxford University Press.
- Dutta, S., & Bhattacharya, S. (2011), “Markov Chain Monte Carlo Based on Deterministic Transformations,”. Submitted, available at <http://arxiv.org/abs/1106.5850>.
- Ghosh, A., Mukhopadhyay, S., Roy, S., & Bhattacharya, S. (2011a), “Bayesian Inference in Nonparametric Dynamic State-Space Models,”. Submitted.
- Ghosh, A., Mukhopadhyay, S., Roy, S., & Bhattacharya, S. (2011b), “Supplement to “Bayesian Inference in Nonparametric Dynamic State-Space Models”,”. Submitted.
- Meinhold, R. J., & Singpurwala, N. D. (1989), “Robustification of Kalman Filter Models,” *Journal of the American Statistical Association*, 84, 479–486.
- Quintana, J., & West, M. (1987), “Multivariate time series analysis: New techniques applied to international exchange rate data,” *The Statistician*, 36, 275–281.
- Santner, T. J., Williams, B. J., & Notz, W. I. (2003), *The design and analysis of computer experiments*, Springer Series in Statistics, New York, Inc.: Springer-Verlag.

- Shumway, R. H., & Stoffer, D. S. (1982), “An Approach to Time Series Smoothing and Forecasting Using the EM Algorithm,” *Journal of Time Series Analysis*, 3, 253–264.
- Shumway, R. H., & Stoffer, D. S. (2011), *Time Series Analysis and Its Applications*, New York: Springer-Verlag.
- Stein, M. L. (1999), *Interpolation of Spatial Data: Some Theory for Kriging*, New York, Inc: Springer-Verlag.
- West, M., & Harrison, P. (1997), *Bayesian Forecasting and Dynamic Models*, New York: Springer-Verlag.